UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXII

COORDINATOR Prof. Claudio Luchinat

Characterization of Intrinsically Disordered Proteins by Nuclear Magnetic Resonance Spectroscopy

Settore Scientifico Disciplinare CHIM/03

PhD student	Tutor
Dott. Maria Grazia Murrali	Prof. Roberta Pierattelli

Coordinator
Prof. Claudio Luchinat

November 2016 – 2019

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Universiteit Utrecht

Acknowledgments

Firstly, I would like to thank my supervisor Prof. *Roberta Pierattelli*, for the trust, the patience and the motivation throughout my doctoral studies. Her guidance helped me in all the time of my PhD journey. I could not have imagined having a better advisor and mentor for my PhD study.

I would like to express my gratitude also to Prof. *Isabella C. Felli*, she mentored me during these three years as much as Prof. Pierattelli, with patient teaching and insightful advices. They both contributed to my growth as a scientist and as a person.

I wish to extend my gratitude to *Alessio Bonucci*, *Letizia Pontoriero, Marco Schiavina* and *Valerio Sainati*, who contributed a lot to my research, creating a wonderful research team.

My sincere thanks to *Fabio Calogiuri, Massimo Lucci, Enrico Morelli, Rebecca Del Conte, Leonardo Gonnelli* and *Marco Allegrozzi* for the technical assistance and the stimulating discussions.

I would like to thank all my colleagues and friends at CERM for the precious help and the great time together.

Finally, special thanks go to my family, for the constant support during this doctorate and during my entire life. Without them, none of this would have been possible.

"Don't Panic!"

D. Adams

Contents

List of abbreviations

AP – Anti-Phase

APSY – Automated Projection Spectroscopy

BEST – Band-Selective Excitation Short-Transient

BMRB – Biological Magnetic Resonance Bank

BSL – Biosecurity Level

BT – BEST TROSY

CBP – CREB Binding Protein

CCR – Cross Correlated relaxation

CREB – Cyclic-AMP-Response-Element-Binding

Cryo–EM – Cryo Electron Microscopy

EDTA – Ethylenediaminetetraacetic acid

E1A – Early region 1A

FID – Free Induction Decay

FT – Fourier Transform

GAPRO – Geometric Analysis Projection algorithm

HeV – Hendra Virus

HSQC – Heteronuclear Single Quantum Correlation

IDP – Intrinsically Disordered Protein

IDR – Intrinsically Disordered Region

IFN – Interferon

IP – In Phase

LCDs – Low Complexity domains

LLPS – Liquid-liquid phase separation

LRE – Longitudinal Relaxation Enhancement

MoRF – Molecular Recognition Feature

MR – Multiple Receivers

NCBD – Nuclear-receptor Coactivator-Binding Domain

NDGA – Nordihydroguaiaretic acid

NiV – Nipah Virus

NMR – Nuclear Magnetic Resonance

NOE – Nuclear Overhauser Effect

NUS – Non-Uniform Sampling

PDB – Protein Data Bank

PNT – P protein N – terminal

PTM – Post-Translational Modification

RDC – Residual Dipolar Coupling

SLiM – Short Linear Motif

SSP – Secondary Structure Propensity

TEM – Transmission Electron Microscopy

TROSY – Transverse Relaxation-Optimized Spectroscopy

Abstract

Intrinsically Disordered Proteins (IDPs) are flexible proteins that challenge structural biology, defeating the structure-function paradigm and requiring tailored methods for their investigation with respect to well-folded proteins. In the last decades, the discovery of their importance in eukaryotic life and their central role in protein interaction networks has prompted their detailed characterization and IDPs are nowadays completely accepted as an important target of research in life sciences.

The increasingly recognized biological relevance of IDPs requires a continuous expansion of the tools for their characterization. Their structural flexibility requires the use of novel experimental methods since X-ray crystallography, which is by far the most utilized technology in structural biology, cannot address these proteins. NMR spectroscopy has been developed into a powerful structural biology technique complementing protein X-ray crystallography and it is the only technology so far able to provide atomic-resolution information of IDPs in their native states.

Exploiting NMR spectroscopy is possible to obtain a trustful description of these dynamic systems, which are able to interconvert through several conformations, providing information under physiological conditions and eventually *in-cell*.

The high flexibility, the little differentiation between the chemical shifts of the nuclei of the same amino acid type and the solvent exposure of the backbone, together with their propensity for aggregation and proteolytic degradation, are only a few of the peculiarities of IDPs that require tailored NMR experiments.

During my doctorate, I focused my studies on the development of NMR methods for the characterization of IDPs and on their application to study challenging cases. The combination of ^{13}C detection with new hardware tools such as Multiple Receivers, and new software tools such as Automated Projection, led to new strategies to speed up NMR experiments. Moreover, modification of base pulse sequences led to simple ^{13}C NMR experiments for the characterization of proline

residues, that are highly abundant in IDPs and less straightforward to be investigated with standard H^N-based NMR experiments.

Together with the development of new methods, I worked on their application to study different aspects of IDPs function. NMR spectroscopy was used for the characterization of a large viral disordered protein, and the study of an unexpected liquid-liquid phase transition likely relevant in the biological mechanism of viral infection. Exploiting NMR we were able to study the interaction between a disordered protein involved in Parkinson's disease and a small molecule that can be a potential drug to contrast the pathology, as well as the interaction between an oncogenic viral protein and the disordered region of a transcriptional coactivator. These interesting applications demonstrate once again the importance of NMR spectroscopy for studying IDPs, adding small but important pieces in the knowledge about the role of protein disorder in biology.

Keywords

NMR, Intrinsically disordered proteins, IDPs, ^{13}C detection, multiple receivers, APSY, LLPS, alpha-synuclein, viral proteins, drug design, Adenovirus, Henipavirus, Nipah virus, CREB binding protein, proline residues.

Chapter 1. Introduction

1.1 Intrinsically disordered proteins

Proteins are chains of amino acids involved in many processes that are crucial for living cells.[1] The dysfunction of these macromolecules is responsible for a large number of pathologies and scientists have been working hard to investigate their structural and functional properties.[2]

For more than a century, the protein structure – function paradigm dominated the scene, stating that protein function is strictly related to a unique well-defined three dimensional structure. After the publication of the first crystal structure of a protein in 1958,[3] the field of structural biology has started to grow rapidly, reaching more than 155000 structures deposited in the Protein Data Bank to date.[4]

Intrinsically disordered proteins (IDPs) are, by definition, a peculiar class of proteins characterized by the absence of a single and stable three-dimensional structure under physiological conditions.[5] Despite disordered and flexible proteins, or intrinsically disordered regions (IDRs), have been identified since long time, only in recent years structural disorder has proven to have an important role in determining protein function, requiring the reassessment of the previous strict paradigm.[5–9]

IDPs are highly abundant in nature, being present in the three domains of life and especially abundant in eukaryotes.[10,11] The essence of disorder is already encoded in the primary sequence, that has peculiar features in disordered proteins.[12] IDPs usually exhibit a low content of hydrophobic residues, thus low hydrophobic driving force for compaction, and high content of charged residues, responsible for strong electrostatic repulsion. Protein disorder is also related to

low sequence complexity, with repetitiveness of series of amino acids.[13,14] In addition, proline residues are largely presents as the lack of the amide proton in their peptide bonds prevents the formation of hydrogen bonds essential to stabilize secondary structural elements such as α-helices and β-strands.[15]

These interesting systems can be also described as dynamic ensembles of multiple conformations separated by low energy barriers, which are therefore able to constantly interconvert.[16] Their high flexibility is essential for their multiple functions.[17] The first intuitive function is as entropic chains in which disordered regions may act as linkers or spacers to give flexibility to larger rigid architecture.[12] IDPs may bind to a target to carry out several functions, altering the binding partner action by activating or repressing its functions. They may act as scavengers by recruiting and storing small ligands,[17] or as assemblers able to stabilize large multiprotein complexes, such as ribosome and chromatin.[18] Interestingly, hub proteins have elevated level of disorder as well, as the flexibility helps promiscuous interactions in one-to-many mode.[9] Recently, IDPs were shown to function as chaperones.[19] Disordered regions in chaperones may bind different partners and its discrimination in binding is mediated by its order-to-disorder transitions. Furthermore, the lack of compaction in IDPs provides accessible sites for post-translational modifications (PTMs), a feature that allows structural and functional regulation (phosphorylation, ubiquitination, acetylation, etc.).[20,21]

To properly describe the interaction strategy of disordered proteins, new binding models compatible with their properties have been developed.[9] Several studies about the interaction of one IDP with a folded partner have been reported, elucidating different mechanisms of interactions. These range from "folding upon binding" mechanism, in which the disordered protein becomes more structured upon interaction, to dynamic complex formation, in which the IDP undergoes rapid exchange between different conformations on the partner surface.[22,23] The latter case is characterized often by rapid switching among multiple interaction sites of the IDP with the target, constituting a so-called *fuzzy complex*.[24] The most intriguing case is the occurrence of interaction between two disordered proteins. This kind of interaction is still poorly understood, although some cases of interaction in which the two disordered proteins maintain their flexibility in a

complex have been very recently proposed.[25–27] Considering the abundance of IDPs in nature, the occurrence of interactions between IDPs could be very frequent. Hence, it is worth to extend our knowledge in this research area.

An important feature in the binding mode of IDPs, is that is driven by the presence of short linear motif (SLiMs), short sequences of 3 to 10 amino acids that can target proteins to a particular subcellular location, regulating low affinity interactions.[28] The flexibility of these SLiMs allows them to adopt various conformations and to bind to multiple partners.[28,29] Their short extension and simple composition causes the high propensity to emerge in unrelated proteins. A consequence of these properties is that pathogenic viruses and bacteria have evolved to mimic these linear motifs to manipulate regulation of cellular processes.[30]

IDPs are highly abundant in eukaryotes[31] and play crucial roles in biological processes; as such their dysfunction is often responsible for important diseases.[32] The most well-known diseases related to IDPs' dysfunction are neurodegenerative diseases like Parkinson's and Alzheimer's diseases.[33] In particular, these two pathologies are caused by misfolding and aggregation of disordered proteins or disordered short regions that can change conformation, leading to the formation of amyloid fibrils that can be accumulated in the afflicted tissue.[34–36]

IDPs' dysfunction is associated with cell cycle deregulation as many of them are involved in regulation, recognition and cellular signalling processes, like for example in the case of the tumour suppressor factor p53 protein.[37] Many other examples of IDPs involved in cancer, diabetes and cardiovascular diseases, have recently been identified through bioinformatics tools for protein disorder predictions.[38]

Recently, many IDPs of viral origin have been characterized, revealing their important role in viral infections.[39] In particular, some viruses are able to interfere with cell cycle regulation by inducing uncontrolled proliferation that may lead to cancer. On the other hand, some viruses with anti-carcinogenic effects have also been identified,[40] increasing the interest for this controversial topic.

1.2 Viral intrinsically disordered proteins

Viruses are obligate intracellular parasites unconditionally dependent on the host cellular machinery for survival and replication.[30] To this end, the virus should be able to interfere with host cell regulation to render the infected cells amenable for replication and expression of the viral proteome. The flexibility of disordered proteins or regions seems to be an essential feature to perform various interactions with several partners at once, allowing a quick adaptation to the host environment.[41] As previously introduced, viruses are able to efficiently alters many regulatory processes by simply mimicking the host SLiMs.[28] SLiMs are ubiquitous in eukaryotic proteomes mediating many distinct processes like targeting proteins to specific subcellular compartments, acting as recognition sites for proteolytic cleavage as well as acting as binding sites in multiple regulatory and signalling roles encoded in hub proteins (Figure 1.1). Thus, the pathogenic mimicry of these motifs gives the opportunity to alter all the regulatory processes of a cell, allowing viruses to hijack and manipulate host proteins for their benefits.[42]
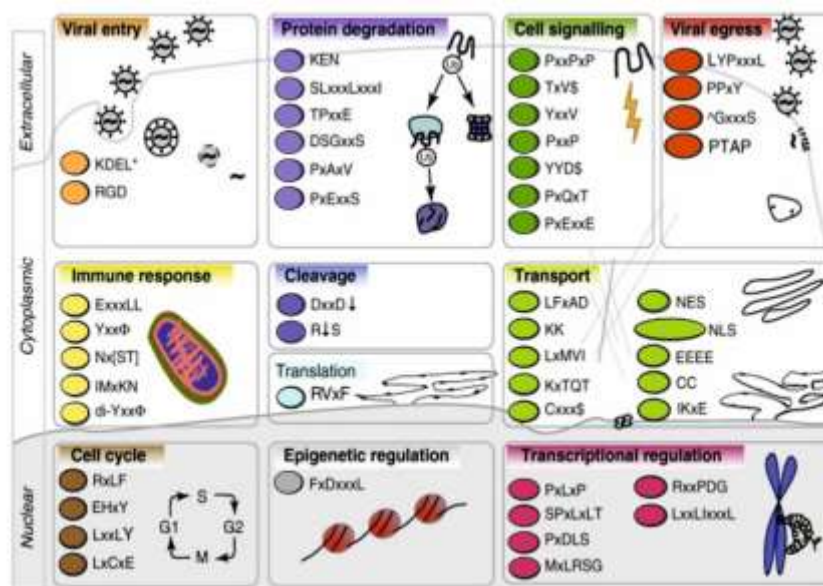
Figure 1.1 *Examples of SLiMs mimicked by various viruses in order to successfully infect the host cell. The motifs are roughly classified by their functions. (Adapted from [30])*

Unfortunately, some viruses have also the capability to trigger uncontrolled proliferation while altering the cellular signalling network, which then ultimately leads to the tumorigenesis.[43] Many viral oncoproteins, like human adenovirus E1A,[44] human papilloma virus E6 and E7[45] and polyomavirus SV40,[46] affect

histone modifications, manipulating the transcription of host genes. In particular, E1A is the most extensively studied viral transcriptional regulator; through its extended intrinsically disordered regions it is able to interact with numerous transcription factors, and the structural characterization of this protein and its interactions constitute an ideal case to elucidate the functional mechanisms of action of oncogenic viruses.[47]

1.3 NMR spectroscopy in structural biology

X-ray crystallography, historically the main technique to access atomic resolution information on large biomolecules, constitutes an excellent tool for the determination of three-dimensional structures of globular proteins and of proteins characterized by a well-defined three dimensional structure in general. A vast number of protein structures were deposited throughout the years in the PDB,[48] with the vast majority of them (89%) being determined through this technology. The success of the structure-function paradigm stimulated the scientific community to focus on protein structure determination and NMR emerged as an alternative tool to enable the 3D structure determination of proteins in solution and, more recently, in the solid state.[49] In the last decade, cryo-EM techniques are emerging for the investigation at near-atomic resolution of protein machineries, flanking X-ray crystallography and NMR spectroscopy.[50]

Once the importance of highly flexible proteins begun to emerge, it was clear the need of designing proper tools for high resolution structural and dynamic investigation of IDPs. Disordered proteins seldom crystalize, and even if they would, the crystals would not provide a complete description of the protein, as the protein would be trapped in only one of the many conformations it is able to sample. Actually, initial evidences of the peculiarity of IDPs was obtained from the investigation of missing patterns in electron density maps, that is of regions of proteins that were not showing up in diffraction because of local flexibility and disorder. Of course X-ray can still be used to determine structures of IDPs that interact with globular proteins through the folding-upon-binding mechanism, but now we know that this is just one of the possible modes of interactions involving IDPs.[51]

In this context NMR constituted one of the key techniques that initially provided experimental support to the growing field of IDPs.[52] Indeed NMR provides information on the structural and dynamic properties of proteins in solution, also in absence of a stable three-dimensional structure. It thus represents the ideal spectroscopic tool to access high resolution information on highly flexible and disordered proteins. It also enables the high resolution investigation of structurally heterogeneous proteins constituted by globular domains separated by flexible linkers and it has become a key tool in modern structural biology.[53]

The first step for the characterization of a protein by NMR is the assignment of all the resonances observed in the NMR spectra to specific nuclear spins in the protein. To this end isotopic enrichment of the biomolecule is generally needed (commonly in ^{15}N and ^{13}C, but also in ^{2}H in particular conditions).[54] Signals chemical shifts are able to provide structural information, being influenced by the local chemical environment of the nuclei. Once the NMR signals are assigned in a sequence-specific manner, several experimental observables can be determined that provide information about the three dimensional structure. The initial analysis of signals chemical shifts can be used to predict elements of secondary structure of the protein, considering their deviation from the random coil values, particularly for C^{α} and C^{β} resonances.[55] Resonance assignment is also necessary to map the interaction of the protein with other proteins or small ligands, providing atomic information of the mode of banding and on the location of the binding site(s).[56]

Most importantly, NMR spectroscopy enables the investigation of the dynamic properties of biomolecules, as it provides information about conformational or chemical exchange, internal mobility and dynamics on timescales ranging from picoseconds to seconds.[57] Longitudinal and transverse ^{15}N relaxation values (relaxation rates R_1 and R_2,) provide information about protein mobility, being sensitive to motions occurring from pico- to nano-seconds timescale. ^{1}H-^{1}H Nuclear Overhauser Effect (NOE), which is a result of cross-talk between different spins in a molecule and depends on the through-space distance between these spins, can be considered the main source of structural information for globular proteins, providing information to calculate secondary structure elements. Other information can be accessible by measuring *J*-coupling constants for dihedral

torsion angles of the protein backbones, by residual dipolar couplings (RDCs) and cross correlated relaxation effects (CCRs) for distance independent projection angles for bond vectors, thus providing information about the spatial arrangements of different domains, or paramagnetic relaxation enhancements, which provide long-range information by deriving distances between the spin label and the NMR active nuclei. Quantification of chemical exchange with the solvent at amide sites reports about compactness and solvent exposure of the protein, which is an important information also for disordered proteins.

During the last fifty years, a wide range of NMR experiments has been developed for the investigation of globular proteins. Despite these experiments provide an excellent starting point for the investigation of large dynamic biomolecules, the peculiarities of IDPs require the development of tailored experiments for their investigation. The rapid interconversion between different conformations, the particular primary sequence with high level of repetition, the solvent exposed backbone, largely affect the dynamic properties of IDPs, influencing the outcome of their NMR spectra. NMR approaches should thus be improved by taking into account the properties of IDPs and their influence on NMR observables in order to exploit at best this very sophisticated spectroscopic technique. The continuous advancement in technologies and instrumentation, provides novel opportunities for methodological developments.

Chapter 2. Novel NMR approaches to characterize IDPs

NMR spectroscopy is the most powerful technique to investigate highly dynamic and flexible systems such as IDPs, provided it is tailored for the characteristics of these proteins. The high flexibility has impact on NMR observables and has to be considered for the design of optimized experiments.

The lack of a well-defined three dimensional structure and the multiple conformations sampled by the polypeptide lead to small differentiation of the chemical shifts of the nuclei of the same amino acid type, averaging most of the chemical shift contributions deriving from the local environment.[58] Thus, NMR spectra of IDPs are characterized by little dispersion of signals in the proton dimension with respect to folded proteins, requiring high-resolution experiments to achieve the sequence specific resonance assignment. The highly redundant amino acid sequence composition of IDPs is also responsible for the presence of large overlapping regions, further complicating the chemical shift assignment. Furthermore, the large abundancy of proline residues, which are not detectable using routine NMR experiments based on $^1H^N$ detection, introduces breaking points in the sequential connection. The use of $^1H^N$–based experiments is further complicated at physiological condition due to the effect of solvent chemical exchange. At high temperature and pH, amide proton – solvent proton exchange reduces the sensitivity of the NMR experiment and causes the broadening of most of the peaks.

The NMR characterization of IDPs has proved to be very challenging. Nowadays, a large number of experiments tailored for IDPs have been designed,[59] but the

development of new methods is always in progress following also the introduction of new technologies and new tools.[60,61]

The small chemical shift dispersion and the extensive overlap of signals in the NMR spectra of IDPs are two major issues in their investigation. Therefore, approaches to obtain increased resolution and to reduce spectral overlaps are essentials. The improvements in instrumentation technology, in particular the introduction of cryogenically cooled probe to obtain higher sensitivity,[62] combined with the use of optimized probehead, enabled the design of new experiments based on heteronuclear direct detection.[63] Study of unfolded proteins demonstrated that ^{13}C and ^{15}N nuclei exhibit greater chemical shift dispersion than proton nuclei ($^{1}H < ^{13}C' < ^{15}N$).[64]

In recent years, ^{13}C direct detection has become a prominent tool for the investigation of IDPs.[66,69,70] In the simplest yet most effective 2D CON experiment, the dispersion of chemical shift is further enhanced exploiting the contributions deriving from the correlation of two neighbouring amino acids. ^{13}C detection represents an excellent compromise between the decreased sensitivity of the ^{13}C nuclei and the increased spectral resolution that is fundamental for IDPs.

Carbon nuclei are not affected by chemical exchange processes with the solvent, providing information about IDPs also at physiological conditions and they are less sensitive to detrimental effect of high salt concentrations in solution, that are often necessary to avoid protein aggregation.[71] Moreover, ^{13}C direct detection provides also an exceptional tool to investigate proline residues which are extensively present in IDPs and yet still less straightforward to be investigated.[72]

2.1 Tailored NMR experiments for detecting proline residues

IDPs have a peculiar primary sequence characterized by high content of charged residues and low content of hydrophobic residues. They are often enriched in so-called disorder promoting amino acids such as glycine, alanine and proline.[5] Despite the abundance of proline residues, they are often poorly investigated in structural and functional studies of IDPs.

Since amide protons are crucial for stabilizing α/β secondary structure elements, the lack of the H^N in proline residues is responsible for being considered promoters of disorder. On the other hand, the rigidity of the cyclic structure of proline residues impart rigidity to the polypeptide, a feature that might be considered in contrast with the high flexibility of IDPs and IDRs. In addition, proline peptide bonds have higher probability of other amino acids to be found in *cis* conformation, further increasing the interest in the characterization of proline-rich proteins.

The lack of the peptidic H^N atoms determine that these amino acids are not detectable using the most frequently used 2D ^1H, ^{15}N NMR experiments. All these controversial aspects render proline residues less straightforward to be investigated, constituting an interesting challenge also by a methodological point of view.

In a C'-N correlation experiment, proline resonances are easily identifiable from all the other signals, having the ^{15}N chemical shift range centred at 137 ppm, far from the other N^H signals, that are significantly up-field. Exploiting this peculiarity, it is possible to introduce a small but very effective modification in 2D CON pulse sequence to select only the proline residues resonances. The use of a band selective 180° pulse on ^{15}N spins of proline residues allows to detect a proline fingerprint with high resolution with a reduced spectral width. The pulse sequence of 2D CON variant employing the ^{15}N band selective pulse to manipulate signals of proline residues spectrum is reported in Figure 2.1.

An interesting application of this experimental scheme is for measuring ^{15}N transverse relaxation rates of proline residues N, which cannot be easily measured using ^1H based experiment and therefore are a missing information in protein characterizations. These can provide information about the local correlation times of different Pro-rich fragments in IDPs or simply about linewidths and coherence lifetimes of ^{15}N spins to optimize experimental parameters set-up. The proline ^{15}N R_2 rates are generally lower than those of non proline residues,
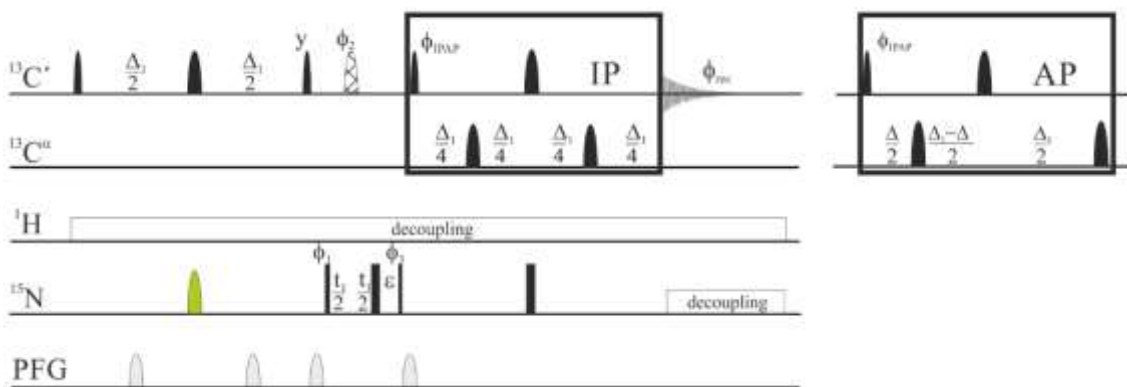
Figure 2.1 *Pulse scheme for the proline-selective 2D CON^{Pro} experiments. The following phase cycling was employed: $\phi1= x, -x; \phi2=2(x), 2(-x); \phi3=4(x), 4(-x); \phi_{rec}= x,-x, x, 2(-x), x, -x, x; \phi_{IPAP}(IP)=x; \phi_{IPAP}(AP)= -y$. The lengths of the delays were: $\Delta_1=1/2\ J(N,C'), \Delta=1/2\ J(C',C\alpha)$. The hatched pulse is an adiabatic pulse that inverts both C' and Cα nuclei, and quadrature detection was achieved through the States time-proportional phase incrementation (States-TPPI) approach by incrementing the phase of the p/2 pulse prior to the evolution period. C' homonuclear decoupling was achieved by using the in-phase/antiphase(IPAP) scheme. The green pulse is the ^{15}N-band selective pulse to invert backbone proline ^{15}N spins only.*

as they lack of the directly bound ^{1}H amide proton and are only influenced by the chemical shift anisotropy contribution to transverse relaxation.

The modified pulse sequence can be designed for 2D CON either using C' and ^{1}H has a starting polarization source, exploiting their different advantages. It is important to remark that this small modification has general applicability. The band selective pulse can be easily introduced in every pulse sequences based on C'-N correlation, providing a simple and straightforward tool for IDPs investigation.

2.1.1 Article: Proline fingerprint in intrinsically disordered proteins

Proline Fingerprint in Intrinsically Disordered Proteins

Maria Grazia Murrali[+],[a] Alessandro Piai[+],[a] Wolfgang Bermel,[b] Isabella C. Felli,*[a] and Roberta Pierattelli*[a]

NMR spectroscopy is one of the main techniques used for high-resolution studies of intrinsically disordered proteins (IDPs), permitting mapping of the structural and dynamic features of all the amino acids constituting the polypeptide at atomic resolution. Only proline residues are less straightforward to characterize because they lack any amide proton, thus rendering them not directly visible in the commonly used 2D ^{1}H,^{15}N correlation experiments. However, proline residues are highly abundant in IDPs and can mediate important functions. In this work we present an easy and effective way to obtain fingerprints of proline residues in IDPs at high resolution.

Intrinsically disordered proteins (IDPs) and regions (IDRs) are often enriched in the so-called disorder-promoting amino acids, among which proline residues have a particular prominence.[1] Despite their high abundance and presence in many atypical motifs characteristic of IDPs/IDRs, the role of proline residues in protein disorder has not yet been addressed in detail.

The many 3D structures available for well-folded globular proteins have shown that the cyclic structure of proline is exploited to increase the rigidity of the protein backbone conformation or to form important structural features, such as tight turns, due to the fact that proline residues' peptide bonds do not contain backbone amide hydrogen atoms, which are crucial for stabilizing α/β secondary structural elements.

Similarly, proline-rich disordered polypeptides largely exploit this amino acid to prevent the formation of stable secondary structures. In terms of dynamic properties, the more constrained backbone conformation arising from the cyclic nature of proline residues, which are in fact used as "rigid spacers" in some techniques,[2] might appear to be in contrast with the high flexibility often observed for IDPs/IDRs, even when highly rich in proline content. Therefore, it is extremely interesting to study the structural and dynamic properties of proline residues in IDPs/IDRs in detail.[3] However, the lack of the peptidic H^N atom means that these amino acids are not directly detectable

in the commonly used 2D ^{1}H,^{15}N NMR spectroscopy and hinders their structural and dynamic characterization by NMR spectroscopy, the only technique able to provide atomic-resolution information on IDPs/IDRs.[4]

Carbon-13 direct-detection NMR has proven to be very valuable for studying IDPs/IDRs.[5–15] In particular, the detection of carbonyl carbon atoms (C') constitutes a particularly useful tool for detection of proline residues.[5,12,16] In this category, the simplest yet most effective experiment is the 2D CON in which the correlations between backbone carbonyl carbon and nitrogen of neighboring residues are detected.[5] The C'–N correlations of proline residues are well separated from all of the other C'–N resonances, their ^{15}N chemical shift range being centered at about 137 ppm whereas the ^{15}N^H envelope of all the other amino acids is significantly upfield. Furthermore, correlating heteronuclear chemical shifts of two consecutive residues (C'$_{i-1}$–N$_i$) helps to resolve signals of different proline residues even in the absence of the contributions to the signal chemical shifts originating from a stable 3D structure.[17]

The basic building block for detecting C'–N correlations in the 2D CON can be further optimized to detect the signals of proline residues selectively and thus to obtain a fingerprint spectrum of these residues. The use of band-selective 180° pulses on ^{15}N spins of proline residues provides a very simple but effective way to detect proline signals selectively by taking advantage of their individual ^{15}N chemical shifts.

The pulse sequence exploiting the ^{15}N-band-selective 180° pulse can be designed in two variants, with either ^{13}C' or ^{1}H$^\alpha$ as a starting polarization source (Figure 1). Both variants allow the study of proline residues near physiological conditions because the nuclei exploited are not subject to chemical-exchange processes with the solvent. The variant exploiting ^{13}C' as a starting polarization source is the most straightforward: sensitivity can be enhanced by exploiting ^{1}H$^\alpha$ as a starting polarization source even if the necessary additional coherence transfer steps increase the length of the pulse sequence. The relatively narrow spectral region in which proline ^{15}N spins resonate also permits sampling of just a small fraction of the ^{15}N dimension, thus reducing the number of increments necessary to achieve a good resolution and the time necessary to measure a 2D NMR map. This aspect can prove particularly useful when fairly low protein concentrations would render quite long the NMR time necessary to acquire the full CON spectrum with sufficient resolution to resolve the proline signals.

In this context, long linkers of large proteins are paradigmatic. As one example, the CREB-binding protein (CBP) is a large (2442 residues) protein in which half of the residues are predicted to form five long intrinsically disordered linkers connecting globular domains.[11,13] Proline residues make up 11.3% of

[a] M. G. Murrali,+ Dr. A. Piai,+ Prof. I. C. Felli, Prof. R. Pierattelli
CERM and Department of Chemistry "Ugo Schiff", University of Florence
Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Florence) (Italy)
E-mail: felli@cerm.unifi.it
        roberta.pierattelli@unifi.it

[b] Dr. W. Bermel
Bruker BioSpin GmbH
Silberstreifen, 76287 Rheinstetten (Germany)

[+] These authors contributed equally to this work.

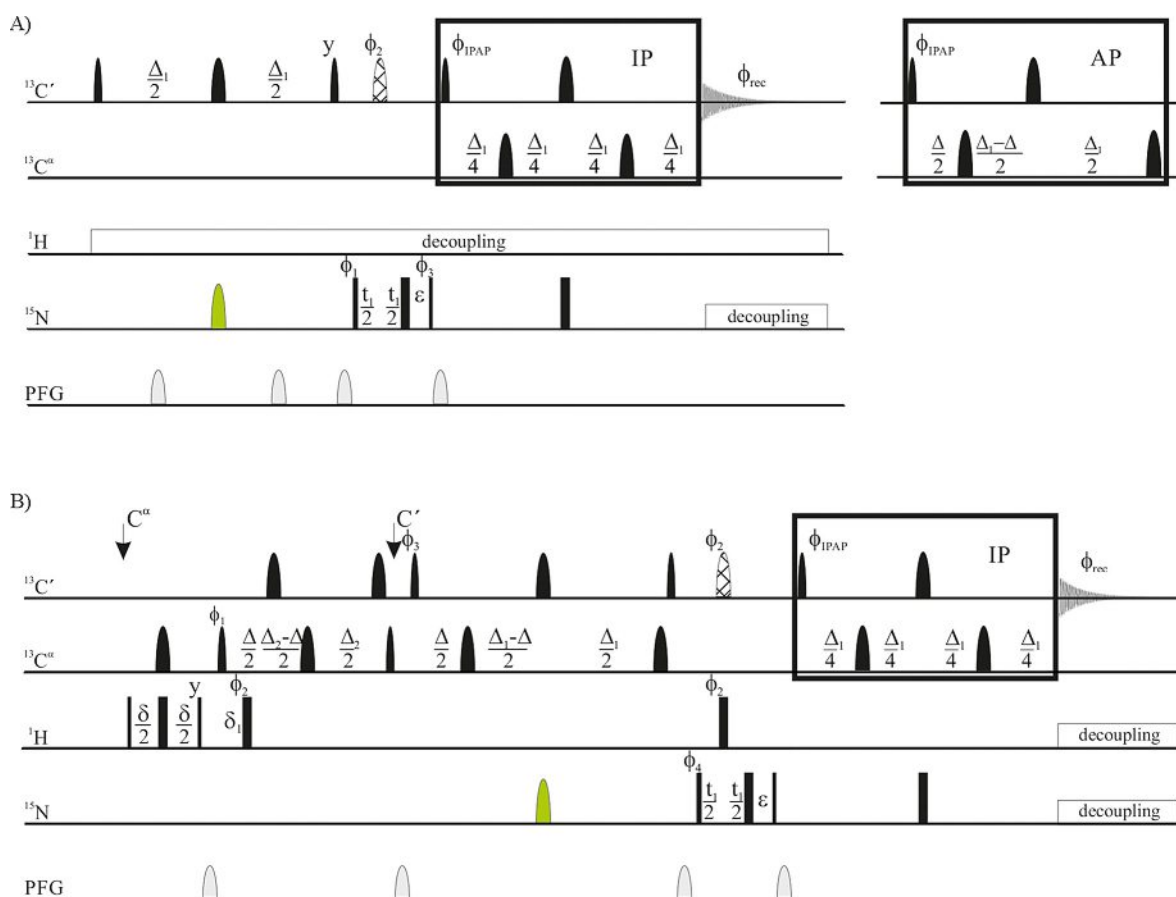Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201800172.

Figure 1. Pulse scheme for the proline-selective A) 2D CON^{Pro}, and B) 2D (H$^\alpha$C$^\alpha$)CON^{Pro} experiments. A) The following phase cycling was employed: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_{rec} = x, -x, x, 2(-x), x, -x, x$; $\phi_{IPAP}(IP) = x$; $\phi_{IPAP}(AP) = -y$. The lengths of the delays were: $\Delta_1 = 1/2\,J(N,C')$, $\Delta = 1/2\,J(C',C^\alpha)$. B) The following phase cycling was employed: $\phi_1 = 4(x), 4(-x)$; $\phi_2 = 8(x), 8(-x)$ $\phi_3 = x, -x$; $\phi_4 = 2(x), 2(-x)$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$; $\phi_{IPAP}(IP) = x$; $\phi_{IPAP}(AP) = -y$. The delays were set as $\delta = 1/2\,J(H^\alpha,C^\alpha)$, $\delta_1 = 1/6\,J(H^\alpha,C^\alpha)$, $\Delta_1 = 1/2\,J(N,C')$, $\Delta = 1/2\,J(C',C^\alpha)$, $\Delta_2 = 1/J(C^\alpha,C^\beta)$. In both sequences, the hatched pulse is an adiabatic pulse that inverts both C' and C$^\alpha$ nuclei, and quadrature detection was achieved through the States time-proportional phase incrementation (States-TPPI) approach by incrementing the phase of the $\pi/2$ pulse prior to the evolution period. C' homonuclear decoupling was achieved by using the in-phase/antiphase (IPAP) scheme. The variant to obtain the AP component is identical in the two experiments and is reported in panel A only. The green pulse is the ^{15}N-band-selective pulse to invert backbone proline ^{15}N spins only.

the total amino acids, and 15.5% of the amino acids in the classified disordered regions.[18] Of the CBP linkers, ID4 is the one containing the highest percentage of proline residues (22% of the total composition) and therefore we used it as a benchmark. It is 207 residues long and contains 45 proline residues, distributed in several little clusters (PPPP, PPP, PP) as well as in several motifs (PXP, PXXP) or sequence repeats of two to three amino acids.[11]

The 2D CON map recorded for ID4 is well-resolved overall (Figure S1 in the Supporting Information), but the 2D CON variant employing the ^{15}N-band-selective pulse to manipulate signals of ^{15}N spins of proline residues (CON^{Pro}) selectively allows a highly resolved 2D fingerprint of the proline region to be acquired in a much shorter time than the complete map (Figure 2). The time saving is essentially proportional to the ratio of the spectral widths necessary to cover the whole ^{15}N dimension with respect to the region in which proline resonances are observed (ca. a factor of 4). It can be seen that a high resolution is very beneficial when the number of proline residues is large, as is often observed in long IDPs/IDRs including

in the present case. The 2D proline fingerprint of ID4 shows that all 45 proline residues can be detected and almost completely resolved, with only two pairs of correlations still in overlap, being located in identical sequence repeats (proline residues 98/103 in PP**P**A and proline residues 132/167 in SM**P**PG). It is interesting to note that the Pro-rich region 92–102, which harbors eight proline residues out of 11 residues in total, instead shows well-resolved signals, an unexpected feature in view of the high redundancy of this part of the polypeptide chain in terms of amino acidic composition. This Pro-rich element precedes the second partially formed helix in ID4 and it is likely to have a functional role that is yet to be explored. Therefore, a possible explanation for the above observation is that the contribution to the chemical shift originating from the transiently structured helix might impact the neighboring proline residues in the cluster 100–102 to a greater extent with respect to those of the cluster 94–97.

Another interesting observation is that the many sequence repeats involving proline residues present in ID4 (VP, LP, SP, AP, TP, QP, MP) are resolved thanks to the excellent resolution of

MQQQIQHRLQ QAQLMRRRMA TMNTRNVPQQ^{30}
SLPSPTSAPP GTPTQQPSTP QTPQPPAQPQ^{60}
PSPVSMSPAG FPSVARTQPP TTVSTGKPTS^{90}
QVPAPPPPAQ PPPAAVEAAR QIEREAQQQQ^{120}
HLYRVNINNS MPPGRTGMGT PGSQMAPVSL^{150}
NVPRPNQVSG PVMPSMPPGQ WQQAPLPQQQ^{180}
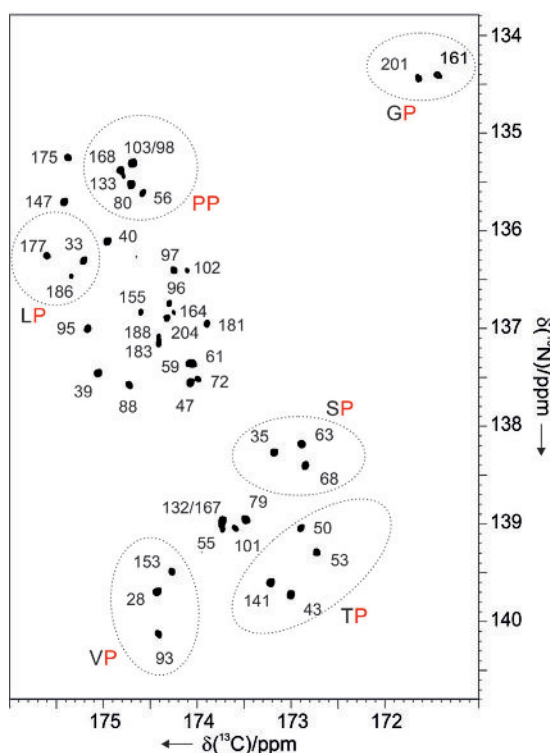PMPGLPRPVI SMQAQAAVAG PRMPSVQ

Figure 2. The 2D proline-fingerprint spectrum of ID4 recorded with use of the 2D CON^{Pro} pulse sequence reported in Figure 1 A. The amino acid sequence of the linker is reported above. In the 2D spectrum the signals are numbered according to the proline position in the sequence; C′–N correlations involving identical residue-type pairs are circled.

the CON^{Pro} experiment (Figure 2). Chemical shift predictors based on the amino acid sequences and taking into account the neighbor effect, such as POTENCI,[19] can in principle predict this distinctive signal distribution. The 2D proline finger-print would provide an easy tool to complete the many IDP

studies performed to date in which assignment statistics are often reported with proline residues excluded,[20,21] on the grounds that only experiments based on H^N detection are used for assignment purposes. Thus, it would be beneficial to increase the statistics about proline chemical shifts in IDPs/IDRs and consequently to further improve the accuracy of chemical shift and secondary structure predictors.

The proline fingerprint is also very useful for the determination of ^{15}N relaxation rates for proline residues, which cannot be easily accessed through the conventional strategy based on ^{1}H detection. These can provide information about the local correlation times of different Pro-rich fragments in IDPs or simply about linewidths and coherence lifetimes of ^{15}N spins to optimize experimental parameters. As an example, the pulse sequence designed to measure the ^{15}N R_2 relaxation rates of proline residues is illustrated in Figure 3. After the creation of ^{15}N in-phase coherence, an additional Carr–Purcell–Meiboom–Gill (CPMG) loop is introduced to determine ^{15}N R_2 relaxation rates. Refocusing pulses on ^{1}H,^{13}C were included to minimize possible development of cross-correlated relaxation.

The proline ^{15}N R_2 rates are reported in Figure 4 and compared with the ^{15}N R_2 data measured for the non-proline residues at the same 16.4 T magnetic field by use of the standard ^{1}H detected experiments[11] and with the secondary structure prediction (SSP) values obtained by measuring the deviation of the experimentally determined chemical shifts for N^H, C′, C^α, and C^β nuclei with respect to the random-coil values.[11] We found the proline ^{15}N R_2 rates to be generally lower than those of non-proline residues, as would be expected from the lack of the directly bound ^{1}H amide proton, but exhibiting a trend that agrees remarkably well with the relaxation profile for non-proline residues, completing the backbone dynamic characterization of ID4. Proline ^{15}N R_2 values confirm that the Pro-rich region encompassing residues 93–103 becomes progressively more structured as it approaches the partially formed helix revealed by the consistently positive values in the SSP plot (residues 101–128, Figure 4 A). Unfortunately, the relaxation rates of the last proline residues of this stretch could not be accurately determined because their signals, although resolved, were considerably less intense than the average for this
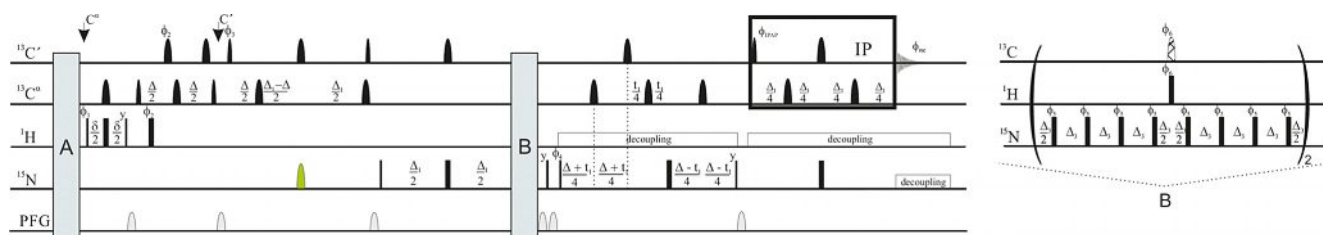

Figure 3. Pulse scheme for 2D (H^{α}C^α)CON^{Pro} providing ^{15}N R_2 values for proline residues. The following phase cycling was employed: $\phi_1 = 4(x)$, $4(-x)$; $\phi_2 = x$; $\phi_3 = x$, $-x$; $\phi_4 = 2(x)$, $2(-x)$; $\phi_5 = 8(x)$, $8(-x)$; $\phi_6 = (-x)$; $\phi_{rec} = x$, $2(-x)$, x, $-x$, $2(x)$, $-x$; $\phi_{IPAP}(IP) = x$; $\phi_{IPAP}(AP) = -y$ (the AP block is as in Figure 1 A). The quadrature detection was achieved through the States-TPPI method by incrementing the phase of the $\pi/2$ pulse prior to the evolution period; C′ homonuclear decoupling was achieved by using the IPAP scheme. The band-selective hatched pulse is an adiabatic pulse that inverts both C′ and C^α nuclei. The lengths of the delays were set as $\delta = 1/2 J(H^α,C^α)$, $\Delta_1 = 1/2 J(N,C')$, $\Delta = 1/2 J(C',C^α)$. The green pulse is the ^{15}N-band-selective pulse to invert backbone proline ^{15}N spins. The ^{15}N relaxation rates were determined by recording several experiments with increasing length of the CPMG loop reported in block B. The expansion shows the minimum block, which was set to 80 ms ($\Delta_3 = 5$ ms), whereas the maximum was set to 800 ms. A block similar to B but not including the ^{1}H pulse (block A) was included at the beginning of the pulse sequence, the length of which was reduced incrementally while the effective CPMG duration was increased, thus keeping the duration of A+B constant and ensuring uniform RF irradiation heating.
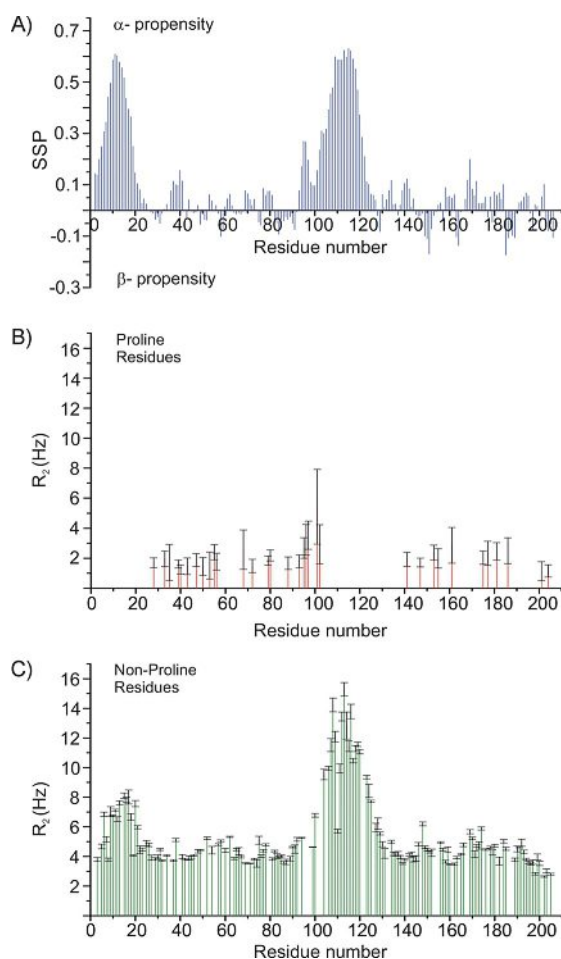
Figure 4. Comparison of A) the SSP data,[11] B) ^{15}N R_2 data measured for proline nitrogen nuclei, and C) ^{15}N R_2 data measured for the non-proline residues,[11] recorded at 16.4 T. Large positive values of the SSP indicate the presence of partially populated helical segments.

sample (ca. 150 μM). However, this observation also points to the fact that the C-terminal part of the poly-P region is more structured than the other proline-rich segments.

The experimental determination of ^{15}N relaxation rates of proline residues also demonstrates that the investigation of these residues in IDPs is particularly promising, because ^{15}N spins, which are among the best-resolved in general, exhibit low ^{15}N transverse relaxation rates; this makes them well suited to achieve very high resolution because their ^{15}N linewidths are significantly narrower than those of all other amino acids. Therefore, a higher resolution can be achieved in any kind of spectrum in which ^{15}N proline spins are sampled, and this observation suggests that they could be used as probes for chemical shift mapping or for interactions and dynamics studies, especially when the ^{1}H detected counterpart experiments suffer from large spectral overlap or line broadening. A similar fingerprint can also be obtained by exploiting ^{15}N direct detection.[22] In principle, ^{15}N detected spectra should provide even better resolved spectra in the direct dimension, due to the intrinsically narrower linewidth of ^{15}N signals. However, ^{15}N detection suffers significantly from the low sensitivity related

to the small ^{15}N gyromagnetic ratio and cannot be widely used for diluted samples such as in the present case.[23]

A comment is due on the general applicability of the simple proline-fingerprint strategy proposed here to the whole suite of C′ detected multidimensional NMR experiments used to study IDPs.[24–28] This very straightforward modification can be easily included in any of the NMR experiments based on C′ detection that exploit the C′–N correlation. For example, this strategy could be used to identify the C^α and C^β resonances both of proline residues and of the preceding amino acids by modifying the 3D CBCANCO[7] and 3D CBCACON[5] experiments. The modification would allow information to be obtained about proline residues that is missing with respect to that available through the analogous H^N detected spectra. The 3D CBCANCO experiment, providing C^γ and C^δ of proline residues, would also allow determination of whether their conformation is *cis* or *trans*. This is an additional example of synergy of different experimental schemes to afford highly complex studies of IDPs. The experiments presented here provide a simple and straightforward NMR tool to contribute to the study of unexplored motifs in IDPs/IDRs.

Experimental Section

The ^{13}C,^{15}N-labeled ID4 protein was prepared as previously described.[11] The NMR experiments were all acquired with ID4 (150 μM) in a water buffer containing phosphate (20 mM) and KCl (100 mM) at pH 6.5 with D$_2$O (10%) added for the lock signal.

All the NMR spectra were acquired with a 16.4 T Bruker AVANCE NEO spectrometer, operating at 700.06 MHz (^{1}H), 176.05 MHz (^{13}C), or 70.97 MHz (^{15}N) and equipped with a cryogenically cooled probe head optimized for ^{13}C direct detection (TXO). Carrier frequencies were placed at 4.7 ppm for ^{1}H, 176.5 ppm or 56.2 ppm for C′ and C^α, respectively, and 137.0 ppm for ^{15}N. Q5 and Q3 shapes of durations of 300 and 231 μs, respectively, were used for ^{13}C band-selective π/2 and π flip angle pulses, except for the adiabatic π pulse to invert both C′ and C^α (smoothed Chirp 500 μs, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength centered at 93 ppm). Decoupling of ^{1}H and ^{15}N was achieved with Waltz (2.6 kHz) and Garp-4 (1.0 kHz) decoupling sequences, respectively. The band-selective 180° pulse on ^{15}N spins of proline residues was achieved with a 8 ms Reburp shape. All gradients employed had a smoothed square shape.

The 2D CON^{Pro} and the (H^{α}C^α)CON^{Pro} experiments were acquired with a spectral width of 29.9 ppm for the ^{13}C dimension and 8.5 ppm for the ^{15}N dimension (1024×128 real points), with 64 scans per increment and relaxation delays of 2.0 s and 1.0 s, respectively. The same parameters were also used for the full 2D CON in Figure S1, with 432 increments and a spectral width of 35.0 ppm for the indirect acquisition dimension.

The 2D (H^{α}C^α)CON^{Pro} experiments to determine the ^{15}N R_2 values for proline residues were acquired with a spectral width of 29.9 ppm for the ^{13}C dimension and 6.5 ppm for the ^{15}N dimension (1024×64 real points), 48 scans per increment, and a relaxation delay of 1.0 s. The following delays were used: 80, 160, 240, 320, 400, 480, 560, 640, 800 ms. The RF field strength used for the CPMG block was 3.1 kHz. The analysis of the data was accomplished by use of the Bruker Dynamic Center 2.4, available as stand-alone ancillary software for TopSpin by Bruker.

In all cases C′ homonuclear virtual decoupling was achieved through the IPAP approach by acquiring both the IP and AP components of the C′—C$^\alpha$ doublet for each increment. These were then subjected to the appropriate linear combination to separate the two multiple components and then shifted to the center of the original doublet to achieve the desired effect.[29]

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] J. Habchi, P. Tompa, S. Longhi, V. N. Uversky, *Chem. Rev.* **2014**, *114*, 6561–6588.
[2] R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18964–18969.
[3] W. J. Wedemeyer, E. Welker, H. A. Scheraga, *Biochemistry* **2002**, *41*, 14637–14644.
[4] *Intrinsically Disordered Proteins Studied by NMR Spectroscopy* (Eds.: I. C. Felli, R. Pierattelli), Springer, Cham, **2015**.
[5] W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, *J. Am. Chem. Soc.* **2006**, *128*, 3918–3919.
[6] Y. Pérez, M. Gairi, M. Pons, P. Bernadò, *J. Mol. Biol.* **2009**, *391*, 136–148.
[7] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J. Magn. Reson.* **2009**, *198*, 275–281.
[8] N. Y. Haba, R. Gross, J. Novacek, H. Shaked, L. Židek, M. Barda-Saad, J. H. Chill, *Biophys. J.* **2013**, *105*, 481–493.
[9] D. Pantoja-Uceda, J. Santoro, *J. Biomol. NMR* **2013**, *57*, 57–63.
[10] D. Sahu, M. Bastidas, S. A. Showalter, *Anal. Biochem.* **2014**, *449*, 17–25.
[11] A. Piai, E. O. Calçada, T. Tarenzi, A. del Grande, M. Varadi, P. Tompa, I. C. Felli, R. Pierattelli, *Biophys. J.* **2016**, *110*, 372–381.
[12] T. Hosek, E. O. Calçada, M. Nogueira, M. Salvi, T. Duarte Pagani, I. C. Felli, R. Pierattelli, *Chem. Eur. J.* **2016**, *22*, 13010–13013.
[13] S. Contreras Martos, A. Piai, S. Kosol, M. Varadi, A. Bekesi, P. Lebrun, A. N. Volkov, K. Gevaert, R. Pierattelli, I. C. Felli, P. Tompa, *Sci. Rep.* **2017**, *7*, 4676.
[14] A. Canales, M. Rosinger, J. Sastre, I. C. Felli, J. Jiménez-Barbero, G. Gimenez-Gallego, C. Fernandez-Tornero, *PLoS One* **2017**, *12*, e0189171.
[15] E. B. Gibbs, F. Lu, B. Portz, M. J. Fisher, B. P. Medellin, T. N. Laremore, Y. J. Zhang, D. S. Gilmour, S. A. Showalter, *Nat. Commun.* **2017**, *8*, 15233.
[16] M. Hornicáková, J. Kohoutová, J. Schlagnitweit, C. Wohlschlager, R. Ettrich, R. Fiala, W. Schoefberger, N. Muller, *Biomol. NMR Assign.* **2011**, *5*, 169–175.
[17] I. C. Felli, R. Pierattelli, *J. Magn. Reson.* **2014**, *241*, 115–125.
[18] H. J. Dyson, P. E. Wright, *J. Biol. Chem.* **2016**, *291*, 6714–6722.
[19] J. T. Nielsen, F. A. A. Mulder, *J. Biomol. NMR* **2018**, *70*, 141–165.
[20] G. Platzer, A. Schedlbauer, A. Chemelli, P. Ozdowy, N. Coudevylle, R. Auer, G. Kontaxis, M. Hartl, A. J. Miles, B. A. Wallace, O. Glatter, K. Bister, R. Konrat, *Biochemistry* **2011**, *50*, 6113–6124.
[21] M. D. Mukrasch, M. von Bergen, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, *J. Biol. Chem.* **2007**, *282*, 12230–12239.
[22] S. Chhabra, P. Fisher, K. Takeuchi, J. J. Ziarek, A. Boeszoermenyi, D. Mathieu, W. Bermel, G. Wagner, H. Arthanari, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1710–1719.
[23] E. B. Gibbs, R. W. Kriwacki, *Methods* **2018**, *138–139*, 39–46.
[24] I. C. Felli, A. Piai, R. Pierattelli, *Ann. Rep. NMR Spectroscop.* **2013**, *80*, 359–418.
[25] W. Bermel, I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, A. Zawadzka-Kazimierczuk, *J. Biomol. NMR* **2013**, *57*, 353–361.
[26] A. Piai, T. Hošek, L. Gonnelli, A. Zawadzka-Kazimierczuk, W. Koźmiński, B. Brutscher, W. Bermel, R. Pierattelli, I. C. Felli, *J. Biomol. NMR* **2014**, *60*, 209–218.
[27] B. Brutscher, I. C. Felli, S. Gil-Caballero, T. Hošek, R. Kümmerle, A. Piai, R. Pierattelli, Z. Sólyom, *Adv. Exp. Med. Biol.* **2015**, *870*, 49–122.
[28] A. Piai, L. Gonnelli, I. C. Felli, R. Pierattelli, K. Kazimierczuk, K. Grudziaz, W. Koźmiński, A. Zawadzka-Kazimierczuk, *J. Biomol. NMR* **2016**, *64*, 239–253.
[29] I. C. Felli, R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2015**, *84*, 1–13.

CHEMBIOCHEM

Supporting Information

Proline Fingerprint in Intrinsically Disordered Proteins

Maria Grazia Murrali+,[a] Alessandro Piai+,[a] Wolfgang Bermel,[b] Isabella C. Felli,*[a] and Roberta Pierattelli*[a]
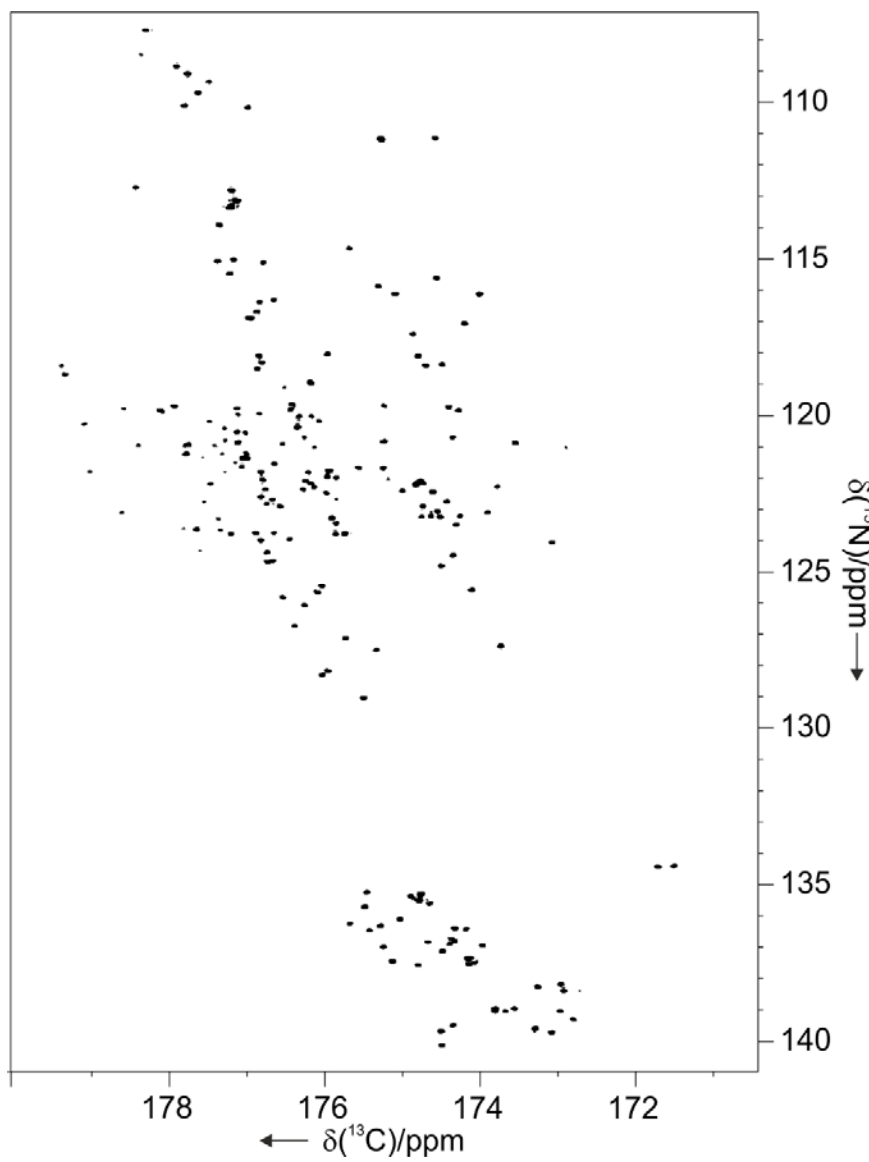
Figure S1. The 2D CON spectrum of CBP-ID4, acquired on a 16.4 T Bruker AVANCE NEO spectrometer at 283K.

2.2 Automated Projection Spectroscopy implementation for ^{13}C – detected NMR

Despite the advantages of using ^{13}C direct detection to increase chemical shifts dispersion, the study of large IDPs requires very often a further push in resolution, which can be obtained with an expansion to more than three dimension experiments. Thanks to fast timescale molecular motions, IDPs have reduced effective rotational correlation times. They present long spin coherence lifetimes (long T_2) and narrow NMR linewidth. Interestingly, the increase in molecular size for IDPs does not affect much relaxation times and linewidths, while it has a prohibitive effect on folded proteins. These relaxation properties allow the design of high dimensional NMR pulse schemes.[73,74]

Increasing the number of indirect dimensions of course will increase the experimental time and complicate the analysis of the data acquired, thus methods based on uniform sampling of indirect dimensions and on conventional processing strategies were not suitable to achieve a high resolution in all dimensions. In recent years, different approaches have been developed based on non-uniform sampling (NUS) to render these experiments feasible in a reasonable amount of time and manageable in terms of processing time and disk space.[75–78]

All these strategies aim to reduce the overall sampled time points by recording only a subset of data points instead of sample points regularly distributed in the time grid as required for all the indirect dimensions in standard NMR experiments. Among the various possible approaches, the Automated projection spectroscopy (APSY-NMR) achieve this goal exploiting the combination of the acquisition of projections of high dimensional NMR experiments. Furthermore, it uses automated peak picking and the subsequent analysis using the geometric analysis projection algorithm (GAPRO).[79–81] The output of an APSY N-dimensional experiment is an N-dimensional peak list with high precision that can be easily analysed manually or by computer algorithm.

Thanks to its implementation if modern software for NMR spectrometers, it is possible to acquire a multidimensional standard NMR experiment from the instruments pulse-sequences library also in APSY mode. Following the flow chart provided by the program, the set-up consists in the selection of the NMR

reference experiment, the selection of magnetization pathway of the experiments that will corresponds to the dimension of the peak list, and the list of angles of projections to be acquired. After the acquisition, the peaks are automatically picked by the GAPRO program resulting in an N-dimensional peak list. The automatic peak picker is not able to distinguish real peaks from spectral artefacts or random noise. Through a series of filter, that can be also user defined but generally suggested by the software, it is possible to select only a list of subgroups from which the peak positions in the N-dimensional space are calculated. The final peak list is the results of combination of redundant experimental data, providing high precision in the peak position. Exploiting high dimensional NMR experiments, up to seven dimensions, it is thus possible to reach a fully automated assignment without any human intervention after the first experimental set up.

While this approach is straightforward for ^{1}H proton detection standard experiments, the application to the whole suite of ^{13}C detection NMR experiments needs a further step in the flow chart to consider the necessity of homonuclear decoupling strategies.[82] After the introduction of the virtual decoupling IP-AP scheme in the APSY format is then possible to exploit this approach also for ^{13}C direct detected experiments.[65,83,84]

Using APSY NMR for the combination of a 5D (HCA)CONCACON and 3D HNCO is thus possible to obtain a fully automated backbone assignment with high precision also for large disordered proteins at physiological condition, exploiting all the advantages of ^{13}C detection previously described and the high resolution provided by multidimensional experiments, in a very reasonable amount of time.

2.2.1 Article: ^{13}C APSY-NMR for sequential assignment of intrinsically disordered proteins

ARTICLE

^{13}C APSY-NMR for sequential assignment of intrinsically disordered proteins

Maria Grazia Murrali[1] · Marco Schiavina[1] · Valerio Sainati[1] · Wolfgang Bermel[2] · Roberta Pierattelli[1,3] · Isabella C. Felli[1,3]

Abstract
The increasingly recognized biological relevance of intrinsically disordered proteins requires a continuous expansion of the tools for their characterization via NMR spectroscopy, the only technique so far able to provide atomic-resolution information on these highly mobile macromolecules. Here we present the implementation of projection spectroscopy in ^{13}C-direct detected NMR experiments to achieve the sequence specific assignment of IDPs. The approach was used to obtain the complete backbone assignment at high temperature of α-synuclein, a paradigmatic intrinsically disordered protein.

Introduction

The functional importance of intrinsically disordered proteins (IDPs) or protein regions (IDPRs) has now been widely recognized (Habchi et al. 2014; Wright and Dyson 2015; van der Lee et al. 2014). Furthermore, a number of key proteins related to the onset of diseases is intrinsically disordered or has large disordered regions (Uversky et al. 2008). Perhaps the most well-known examples include proteins involved in the development of neurodegenerative diseases, such as α-synuclein and Tau for Parkinson's and Alzheimer's diseases. In addition, a vast range of examples of IDPs/IDPRs related to cancer progression are emerging, beyond the most

✉ Roberta Pierattelli
pierattelli@cerm.unifi.it

✉ Isabella C. Felli
felli@cerm.unifi.it

1 CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

2 Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

3 Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino, Florence, Italy

well-known example of p53. These include c-myc, p21, AR, BRCA to name just a few recent examples (Uversky et al. 2014). Therefore, novel approaches to target IDPs or IDPRs are urgently needed in the field of drug discovery (Tóth et al. 2014; Ambadipudi and Zweckstetter 2016; Joshi et al. 2016; Heller et al. 2017) and in this frame it becomes very important to develop NMR methods to study IDPs/IDPRs in their native state under physiological conditions (Felli and Pierattelli 2015b).

The peculiar amino acidic composition of IDPs, often characterized by multiple sequence repetitions and low complexity regions, the high flexibility typical of IDPs/IDPRs and the many conformations sampled in which the protein backbone is largely solvent exposed have an impact on NMR observables that should be considered in the design of effective experimental NMR approaches (Brutscher et al. 2015).

The critical points that need to be faced in order to optimize NMR spectra for the study of highly flexible IDPs/IDPRs include the very low chemical shift dispersion typical of proteins that lack a stable, well defined 3D structure, as well as solvent exchange processes that broaden amide proton resonances when approaching physiological conditions. As known since the early studies on protein folding, heteronuclear spins are characterized by a large chemical shift dispersion and thus well suited to characterize disordered proteins (Dyson and Wright 2001). Indeed, initial studies of unfolded proteins were stimulated by the development of 3D triple resonance experiments

(Kay et al. 1990; Sattler et al. 1999). However, to study unfolded/intrinsically disordered proteins of increasing complexity, a further push in resolution was necessary; a major contribution towards this goal derived from the extension of the dimensionality of NMR experiments to more than three dimensions. Of course methods based on uniform sampling of indirect dimensions and on conventional processing strategies were not suitable to achieve a high resolution in all dimensions, a key aspect to study complex IDPs/IDPRs; different approaches needed to be developed to render these experiments feasible in a reasonable amount of time and manageable in terms of processing time and disk space. Many non-uniform sampling approaches and different strategies to process the data were indeed proposed (Kupce and Freeman 2004; Kim and Szyperski 2003; Hiller et al. 2005; Hoch and Stern 2001; Orekhov et al. 2003; Kazimierczuk et al. 2006) and initially implemented in triple resonance multidimensional experiments based on H^N detection (Narayanan et al. 2010; Fiorito et al. 2006; Zawadzka-Kazimierczuk et al. 2010; Motackova et al. 2010) and on H^α detection (Yao et al. 2014; Mäntylahti et al. 2011) showing a great potential for the characterization of IDPs (Fiorito et al. 2006; Zawadzka-Kazimierczuk et al. 2012; Narayanan et al. 2010; Yao et al. 2014).

In parallel, the increase of the sensitivity of NMR instrumentation (Kovacs et al. 2005) enabled the development of experimental strategies based on heteronuclear direct detection (Bermel et al. 2006c), which turned out to be an excellent tool to study IDPs. Indeed the valuable chemical shift dispersion of heteronuclear spins can be exploited in all dimensions of NMR experiments, including the directly detected one, and limitations deriving from solvent exchange broadening are obviously reduced. Several methods to overcome the problem of ^{13}C homonuclear couplings in the direct acquisition dimension were proposed (Felli and Pierattelli 2015a, b) and enabled the application of C′ direct detection with high resolution. A suite of 3D multidimensional C′ detected experiments was proposed (Bermel et al. 2006b, 2008, 2009a, b; Felli et al. 2009; O'Hare et al. 2009) and proved to be very useful for sequence specific assignment and characterization of several IDPs. The suite was then expanded to include also C′ detected experiments of dimensionality higher than 3 by implementing random non-uniform sampling in combination with sparse multidimensional Fourier transform (SMFT) processing of the data showing that experiments of higher dimensionality based on C′ detection are indeed possible and constitute a valuable tool to study IDPs (Novácek et al. 2011; Bermel et al. 2012, 2013; Novácek et al. 2012; Haba et al. 2013; Piai et al. 2014; Baias et al. 2017).

As a further step facilitating the exploitation of the ^{13}C-detection based assignment strategy, we present the implementation of projection spectroscopy for C′ detected multidimensional NMR experiments. The approach is tested on α-synuclein at high temperature.

Materials and methods

Isotopically enriched α-synuclein was prepared as previously described (Huang et al. 2005). A sample of 1.0 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 200 mM NaCl, 0.5 mM EDTA, 20 mM phosphate buffer at pH 6.5 was used. 10% D_2O was added for the lock.

All the 2D NMR experiments were acquired at 285.5, 295, 305 and 315 K while the APSY experiments were all acquired at 315 K on a Bruker AVANCE NEO spectrometer operating at 700.06 MHz ^1H, 176.05 MHz ^{13}C and 70.97 MHz ^{15}N frequencies equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection (TXO). Carrier frequencies and RF pulses suitable for triple resonance ^{13}C detected experiments were used and are summarized hereafter while specific parameters for the different experiments are reported in Table 1. Q5 and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 300 and 231 μs, respectively, were used for ^{13}C band-selective π/2 and π flip angle pulses, except for the π pulses that should be band-selective on the C^α region (Q3, 1200 μs) and for the adiabatic π pulse to invert both C′ and C^α (smoothed Chirp 500 μs, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength) (Böhlen and Bodenhausen 1993). The ^{13}C band selective pulses on C^α and C′ were applied at the center of each region, 53 and 173.5 ppm respectively; the Chirp pulse was centered at 90 ppm. Carrier frequencies for ^{15}N and ^1H were 122.5 and 4.7 ppm respectively. Decoupling of ^1H and ^{15}N was achieved with Waltz (2.6 kHz) and Garp-4 (1.0 kHz) decoupling sequences respectively (Shaka et al. 1985). All gradients employed had a smoothed square shape. The parameters used for the acquisition of the 5D and 4D experiments as well as the parameters selected to implement the APSY approach are reported in Tables 1 and 2 respectively. FLOPSY-16 (Kadkhodaie et al. 1991) was used in the 4D experiment to spin-lock. All the spectra were acquired using Bruker TopSpin 4.0.1 software. Calibration of the spectra was achieved using DSS as a standard for ^1H and ^{13}C; ^{15}N shifts were calibrated indirectly (Markley et al. 1998).

Results and discussion

The 2D ^1H–^{15}N HSQC and 2D ^{13}C–^{15}N CON experiments of α-synuclein recorded at various temperatures are reported in Fig. 1. It can be noted that the number of cross peaks that can be observed through 2D ^1H–^{15}N HSQC spectra

Table 1 Acquisition parameters for 5D^{APSY}-(HCA)CONCACON (Bermel et al. 2013b) and 4D^{APSY}-HCCCON experiments (Bermel et al. 2012)

Experiment	Spectral width (Hz) and maximal evolution times				NS	Inter scan delay (s)	Dimension of acquired data					Duration of the experiment	Number of projections
	F1	F2/F4	F3	F5			F1	F2	F3	F4	F5		
5D (HCA)CONCACON	1695 (^{13}C') 28.4 ms	2404 (^{15}N) 53.2 ms	4505 (^{13}C$^{\alpha}$) 75.5 ms	9090 (^{13}C') 112.6 ms	8	0.8	256	256	256	256	2048	4 days 18 h	64
	F1	F2	F3	F4			F1	F2	F3	F4	F5		
4D HCCCON	3840 (^{1}H) 66.6 ms	10,870 (^{13}C^{ali}) 23.6 ms	2404 (^{15}N) 106.5 ms	9090 (^{13}C') 112.6 ms	4	0.8	512	512	512	2048	/	2 days 23 h	41

decreases with increasing temperature. This is the result of the pronounced exchange processes of amide protons with water protons due to the largely solvent exposed backbone of α-synuclein.

Nuclear spins of non-exchangeable atoms should thus be exploited in NMR experiments. However, even if in principle very sensitive, ^{1}H nuclear spins of aliphatic and aromatic residues are characterized by a quite narrow chemical shift dispersion, in particular when considering amino acids of the same (or similar) type in proteins lacking a 3D structure such as in IDPs. In addition the ^{1}H–^{1}H homonuclear couplings, which are not easy to remove in the direct acquisition dimension, constitute an additional drawback when really high resolution is needed. Instead, non-exchangeable heteronuclear spins, such as ^{13}C and ^{15}N, provide a large chemical shift dispersion even if at the expense of reduced sensitivity when directly detected. Several strategies can be implemented for C' homonuclear decoupling (Felli and Pierattelli 2015a). Thus, correlations between heteronuclear spins involved in the peptide bond are the most appropriate ones to achieve high resolution (Bermel et al. 2013a), as evidenced in the spectra reported in Fig. 1 which could be acquired with just a few scans per increment; 2D ^{13}C–^{15}N CON spectra (Bermel et al. 2006a) thus provide a valuable tool to study IDPs at physiological conditions (Gil et al. 2013).

The 2D ^{13}C–^{15}N CON spectra can then be expanded into higher dimensional experiments that provide different types of correlations between different nuclear spins in order to achieve sufficient information for sequence specific assignment. Increasing the number of dimensions however, even when exploiting heteronuclear spins to take advantage of their high dispersion in IDPs, is effective only if we are able to achieve a very high resolution in the indirectly sampled dimensions. This is practically not feasible due to time constraints as well as disk usage unless non-uniform sampling strategies are used, which, combined with appropriate processing strategies, have shown to be very effective (Bermel et al. 2012, 2013b; Novacek et al. 2011, 2012). Here we test the applicability of projection spectroscopy and of the relative automated data analysis (Hiller et al. 2005) to the C' detected 5D (HCA)CONCACON experiment (Bermel et al. 2013b), which correlates the CON peak of one peptide bond with the preceding one, for the assignment of α-synuclein at high temperature (315 K).

The four orthogonal planes that can be collected through this experiment, reported in Fig. 2, show the excellent resolution that can be achieved in all dimensions thanks to the contribution of heteronuclear chemical shifts, in particular of C' and N that provide correlations across one peptide bond (A) and of C' and N belonging to the same amino acid (B), as well as sequential C'–C' correlations (C). The plane correlating C' and C$^{\alpha}$ (D) is characterized by a somehow

Table 2 Values of the projection angles α, β, γ, and of the spectral widths in the dimensions ω_{1-4} used here for recording the 2D-projections of the 5D^{APSY}-(HCA)CONCACON experiment. The resulting linear combinations of frequencies are given in the right column

α	β	γ	Spectral width (Hz)	Linear combination
0°	0°	0°	2404	ω_4
0°	0°	90°	1695	ω_1
0°	90°	0°	2404	ω_2
90°	0°	0°	4505	ω_3
±28°	0°	0°	4238	$\omega_4\cos(28°) \pm \omega_3\sin(28°)$
0°	±45°	0°	3400	$\omega_4\cos(45°) \pm \omega_2\sin(45°)$
0°	0°	±55°	2767	$\omega_4\cos(55°) \pm \omega_1\sin(55°)$
90°	±62°	0°	4238	$\omega_3\cos(62°) \pm \omega_2\sin(62°)$
90°	0°	±69°	3197	$\omega_3\cos(69°) \pm \omega_1\sin(69°)$
0°	90°	±55°	2767	$\omega_2\cos(55°) \pm \omega_1\sin(55°)$
±28°	±41°	0	4775	$\omega_4\cos(28°)\cos(41°) \pm \omega_2\sin(41°) \pm \omega_3\sin(28°)\cos(41°)$
±28°	0°	±51°	3984	$\omega_4\cos(28°)\cos(51°) \pm \omega_1\sin(51°) \pm \omega_3\sin(28°)\cos(51°)$
0°	±45°	±45°	3603	$\omega_4\cos(45°)\cos(45°) \pm \omega_1\sin(45°) \pm \omega_2\sin(45°)\cos(45°)$
90°	±62°	±51°	3984	$\omega_3\cos(62°)\cos(51°) \pm \omega_1\sin(51°) \pm \omega_2\sin(62°)\cos(51°)$
±15°	0°	0°	3488	$\omega_4\cos(15°) \pm \omega_3\sin(15°)$
0°	±27°	0°	3233	$\omega_4\cos(27°) \pm \omega_2\sin(27°)$
0°	0°	±35°	2941	$\omega_4\cos(35°) \pm \omega_1\sin(35°)$
90°	±43°	0°	4934	$\omega_3\cos(43°) \pm \omega_2\cos(43°)$
90°	0°	±53°	4065	$\omega_3\cos(53°) \pm \omega_1\sin(53°)$
0°	90°	±35°	2941	$\omega_2\cos(35°) \pm \omega_1\sin(35°)$
±47°	0°	0°	4934	$\omega_4\cos(47°) \pm \omega_3\sin(47°)$
0°	±63°	0°	3233	$\omega_4\cos(63°) \pm \omega_2\sin(63°)$
0°	0°	±71°	2385	$\omega_4\cos(71°) \pm \omega_1\sin(71°)$
90°	±75°	0°	3488	$\omega_3\cos(75°) \pm \omega_2\cos(75°)$
90°	0°	±79°	2523	$\omega_3\cos(79°) \pm \omega_1\sin(79°)$
0°	90°	±71°	2385	$\omega_2\cos(71°) \pm \omega_1\sin(71°)$
±15°	0°	0°	3488	$\omega_4\cos(15°) \pm \omega_3\sin(15°)$
0°	±20°	0°	3081	$\omega_4\cos(20°) \pm \omega_2\sin(20°)$
±70°	0°	0°	5056	$\omega_4\cos(70°) \pm \omega_3\sin(70°)$
0°	±70°	0°	3081	$\omega_4\cos(70°) \pm \omega_2\sin(70°)$

reduced resolution but still provides useful information about the amino acid type. These are excellent requisites to implement projection spectroscopy through which not only orthogonal planes but also "transverse" planes cutting the multidimensional object at different angles are acquired with very high resolution as shown in panels E and F of Fig. 2. Thanks to the excellent resolution that can be obtained in the projections of C′ detected exclusively heteronuclear experiments, it is possible to resolve the majority of cross peaks in the projections and to identify essentially all correlations expected in the 5D spectrum just by inspection of a limited set of projections. In the present case, 64 projections were acquired with very high resolution (256 points were set in each of the 4 indirect dimensions); it is worth noting that to achieve this kind of resolution for the analogous 5D experiment acquired with uniform sampling would have required a prohibitively long time (hundreds of years!). The analysis of the 64 projections, which can be performed automatically by

using the algorithm GAPRO (Geometric Analysis of PROjections) (Hiller et al. 2008; Hiller and Wider 2012) as it is implemented in the TopSpin 4.0.1 software, provides 191 5D correlations (131 C′$_{i-1}$, N$_i$, C$^{\alpha}_i$, C′$_i$, N$_{i+1}$ and 60 C′$_{i-1}$, N$_i$, C$^{\alpha}_{i-1}$, C′$_{i-1}$, N$_i$). These correlations constitute an excellent tool to achieve sequence specific assignment by linking the shifts of C′ and N involved in a peptide bond with those of the previous one, and also giving some hint on the kind of amino acid through the information of the C$^{\alpha}$ chemical shift. The correlations identified through the GAPRO analysis of the projections are stored in the form of N-dimensional peaklists that can be manually inspected by loading them on the different projections. In particular, by loading the peak-lists on a 2D ^{13}C–^{15}N CON spectrum the completeness of the assignment can be verified and it is possible to perform the sequence specific walk through the backbone just by moving from one cross peak (C′$_i$, N$_{i+1}$) to the neighboring one (C′$_{i-1}$, N$_i$), as shown schematically in Fig. 3 for a few residues. It
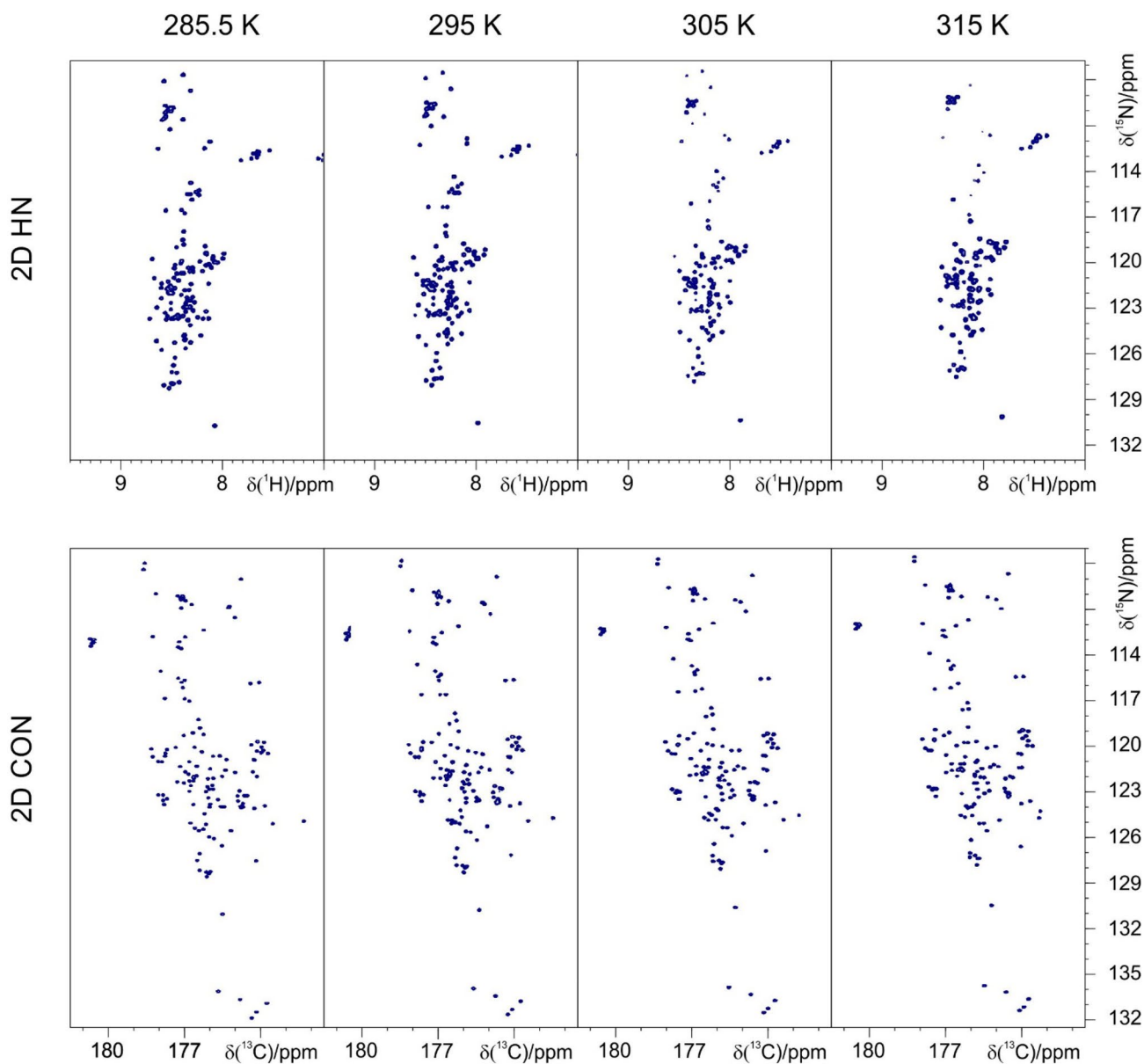
Fig. 1 ^{1}H–^{15}N HSQC (top) and ^{13}C–^{15}N CON-IPAP spectra (bottom), acquired at different temperatures (285.5, 295, 305, 315 K), showing the temperature dependence for α-synuclein signals. Experiments were acquired at 16.4 T on α-synuclein at pH 6.5

is worth noting that for most of these residues the H^N resonance was not observable in ^{1}H detected experiments.

A few comments are due on technical details related to the implementation of the APSY approach for the 5D (HCA)CONCACON experiment (Bermel et al. 2013b). In contrast to the original sequence, the t_2 period was implemented in a semi-constant time manner in order to increase the resolution in this dimension. As noted above, the version presently used yields two types of correlations: type I (C′$_{i-1}$, N$_i$, C^α_i, C′$_i$, N$_{i+1}$) which contain the information for sequential assignment and type II (C′$_{i-1}$, N$_i$, C$^α_{i-1}$, C′$_{i-1}$, N$_i$) which correlate three nuclear spins (C$^α_{i-1}$, C′$_{i-1}$, N$_i$) giving

essentially the same information that can be found in 3D CACON experiments. The experiment is optimized to detect correlations of type I; correlations of type II can in principle be suppressed as suggested for the 3D (HCA)CON(CA)H (Mäntylahti et al. 2011). However, in the present case this subset of peaks could be easily identified directly from the peak list and further used for confirming the assignment, so that no further modification of the experiment was required. A few type I correlations were instead missing. Manual inspection of the peak list revealed that most of them involve residues in the known α-synuclein repeats which could not all be safely identified through the automatic algorithm.
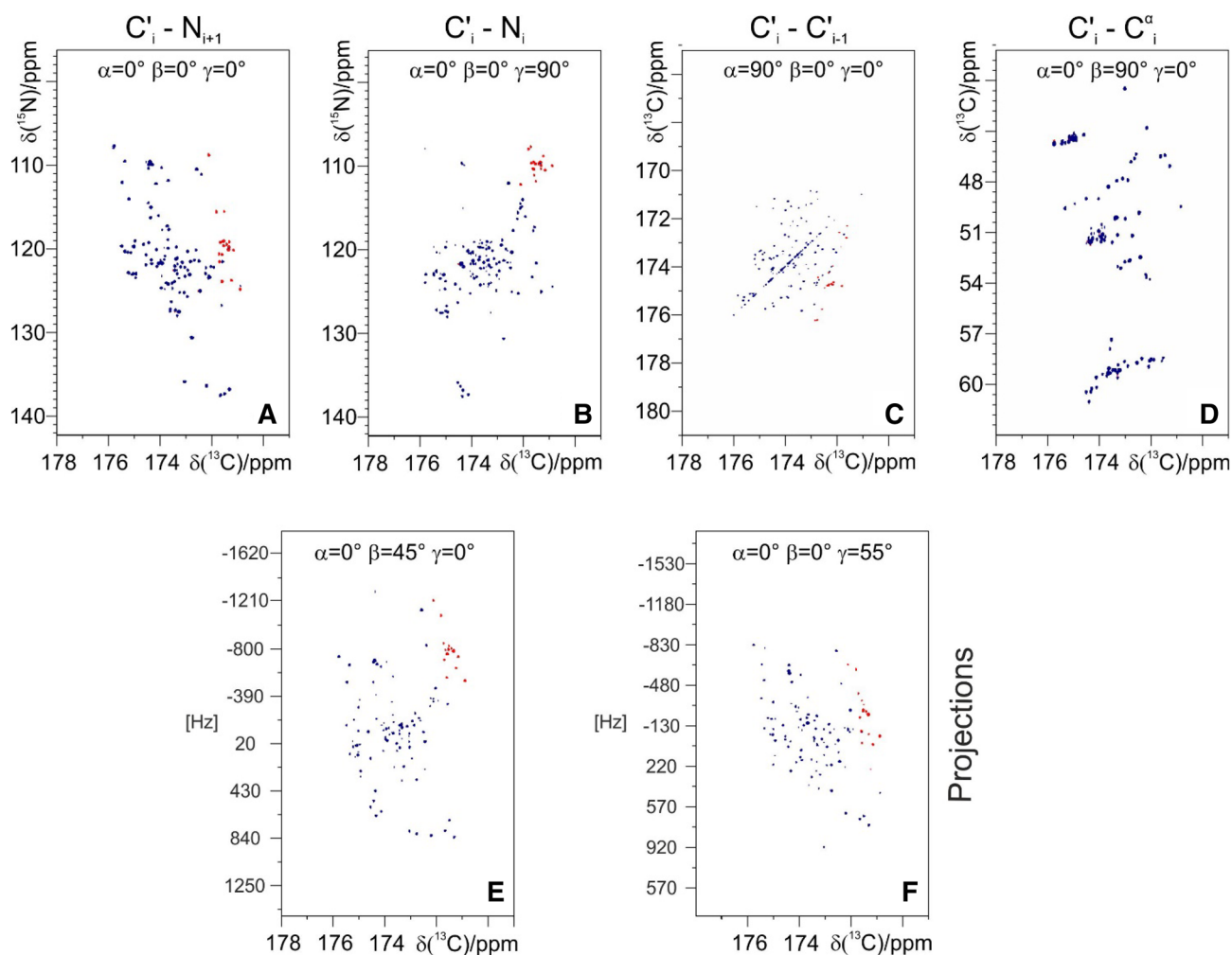
Fig. 2 A subset of the 2D projections acquired with the 5D (HCA) CONCACON experiment are reported to show the quality of the data and to provide an intuitive picture of how the APSY approach works with C′ detected experiments on IDPs. Orthogonal projections (**A–D**) and non-orthogonal ones (**E–F**) are shown. The spectra were recorded at 16.4 T on α-synuclein at 315 K. The four orthogonal projections correspond to C′(i)–N(i+1), C′(i)–N(i), C′(i)–C′(i±1), and Cα(i)–C′(i). In the two non-orthogonal projections the y-coordinate units are arbitrary, as the frequencies in the indirect dimensions are a linear combination of frequencies, that is (E) $\omega_2\sin(45°)$ + $\omega_4\cos(45°)$ and (F) $\omega_1\sin(55°)$ + $\omega_4\cos(55°)$

A second comment concerns the number of projections that need to be acquired to obtain the information for sequence specific assignment. In the present case we collected projections at 30 different combinations of angles α, β, γ that, considering the issue of quadrature detection in all indirect dimensions, translates into 64 different 2D projections (Table 2). However, processing of the data using a smaller number of projection angles shows that the information content is essentially the same if we use only 26 combinations of angles, as suggested by the software (Hiller and Wider 2012; Hiller et al. 2008), and that the number of type I correlations detected in the spectra is only slightly reduced if we consider a smaller subset of angles (Table 3), showing the robustness of this approach. The major issue to be solved was to properly implement in the APSY software the virtual

homonuclear decoupling of C′ through the IPAP approach. The processing of the data, including the virtual decoupling, can now be handled very easily through TopSpin 4.0.1.

The APSY approach can of course be implemented in any kind of C′ detected experiment, using the 2D ^{13}C–^{15}N CON spectrum as a reference. For example, the information about the amino acid side chain, can be easily obtained and correlated to each of the cross peaks observed in the 2D ^{13}C–^{15}N CON spectrum, by collecting experiments such as 4D HCBCACON (Bermel et al. 2009a, b) or 4D HCCCON (Bermel et al. 2012). As an example the latter was acquired on α-synuclein and provided information about the vast majority of ^{13}C chemical shifts of aliphatic side chains. Care was taken to implement the ^{13}C evolution in the constant-time mode in order
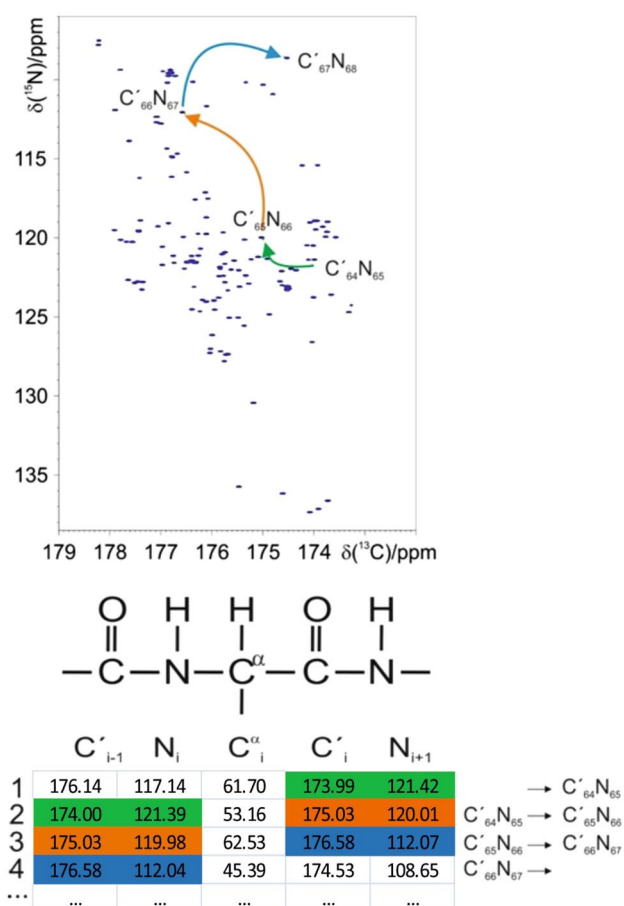
Fig. 3 As an example the assignment strategy using the peak lists resulting from the GAPRO analysis of the 5D (HCA)CONCACON projections is shown for residues 65–67 of α-synuclein. Four five-dimensional "type I" correlations (C'_{i-1}, N_i, C^{α}_i, C'_i, N_{i+1}) are shown. It is worth noting that the amide proton resonances of these residues were not detectable

Table 3 Analysis of the 5D^{APSY}-(HCA)CONCACON processed considering a reduced number of projection angles as described in the text

Angle sets	Projections	Experimental time	Identified correlations (Type I)
30	64	4 days, 18 h	131
26	56	4 days, 5 h	131
20	44	3 days, 9 h	129
10	16	1 day, 2 h	116

to refocus homonuclear ^{13}C couplings of aliphatic carbons that easily show up in highly resolved projections and can complicate the analysis of crowded regions. The modified pulse sequence (4D CT-HCCCON) and a few projections to illustrate the quality of the spectra are reported in the Supplementary Material (Figs. S1 and S2).

A final comment is due about the amide proton resonances that are still observable approaching physiological conditions. Combination of the C′ detected experiments with H^N detected ones provides a useful tool to achieve unambiguous assignment of the amide proton resonances that are still observable in these conditions. In our case we collected a 3D HNCO experiment, also in the APSY mode, in order to correlate H^N_i resonances with the C'_{i-1}–N_i resonances that we have assigned through C′ detected experiments. Through this approach 93 cross peaks were detected and, thanks to the peak list obtained in a fully automated manner through the GAPRO analysis of the APSY HNCO, the available assignment of heteronuclear spins could easily be extended to the residual H^N resonances that can be detected at this temperature. It is worth noting that a similar result could be achieved by following the temperature dependence of the signals. The latter approach however works well only for well resolved H^N cross peaks but it becomes quite ambiguous for cross peaks in crowded regions of the 2D ^{1}H–^{15}N HSQC spectra. Therefore the combination of C′ detected experiments with H^N detected ones, even at high temperature and pH, provides unambiguous information to assign H^N resonances that remain also in presence of fast exchange, without ambiguities resulting from temperature dependence of the signals and/or from incomplete information provided by H^N detected spectra. The temperature dependence of H^N, C′ and N spins, once unambiguous assignment is available, might provide also valuable information to investigate variations of exchange processes of amide protons with the solvent upon increasing temperature as well as to highlight possible variations in partially populated local conformations by analyzing heteronuclear chemical shifts. The sequence specific assignment of α-synuclein achieved at 315 K has been deposited in the BMRB (ID 27348).

Concluding, we demonstrate the implementation of the APSY strategy in C′ detected multidimensional NMR experiments. Excellent resolution can be achieved for the 2D cross sections of the high dimensional C′ detected exclusively heteronuclear spectra, showing that the two approaches (APSY and C′ detection) mutually benefit by their joint implementation.

References

Ambadipudi S, Zweckstetter M (2016) Targeting intrinsically disordered proteins in rational drug discovery. Expert Opin Drug Discov 11:65–77

Baias M, Smith PE, Shen K, Joachimiak LA, Zerko S, Koźmiński W, Frydman J, Frydman L (2017) Structure and dynamics of the Huntingtin exon-1 N-terminus: a solution NMR perspective. J Am Chem Soc 139:1168–1176

Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006a) Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64

Bermel W, Bertini I, Felli IC, Lee Y-M, Luchinat C, Pierattelli R (2006b) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128:3918–3919

Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006c) ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

Bermel W, Felli IC, Kümmerle R, Pierattelli R (2008) ^{13}C direct-detection biomolecular NMR. Concepts Magn Reson 32A:183–200

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009a) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bermel W, Bertini I, Felli IC, Pierattelli R (2009b) Speeding up ^{13}C direct detection biomolecular NMR experiments. J Am Chem Soc 131:15339–15345

Bermel W, Bertini I, Gonnelli L, Felli IC, Koźmiński W, Piai A, Pierattelli R, Stanek J (2012) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S (2013a) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR 55:231–237

Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A (2013b) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361

Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

Brutscher B, Felli IC, Gil-Caballero S, Hošek T, Kümmerle R, Piai A, Pierattelli R, Sólyom Z (2015) NMR methods for the study of instrinsically disordered proteins structure, dynamics, and interactions: general overview and practical guidelines. Adv Exp Med Biol 870:122

Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–271

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

Felli IC, Pierattelli R (2015a) Spin-state-selctive methods in solution- and solid-state biomolecular ^{13}C NMR. Prog NMR Spectrosc 84:1–13

Felli IC, Pierattelli R (eds) (2015b) Intrisically disordered proteins studied by NMR spectroscopy. Springer, Switzerland

Felli IC, Pierattelli R, Glaser SJ, Luy B (2009) Relaxation-optimised Hartmann-Hahn transfer for carbonyl-carbonyl correlation spectroscopy using a specifically tailored MOCCA-XY16 mixing sequence for protonless ^{13}C direct detection experiments. J Biomol NMR 43:187–196

Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR 35:27–37

Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR studies of intrinsically disordered proteins near physiological conditions. Angew Chem Int Ed 52:11808–11812

Haba NY, Gross R, Nováček J, Shaked H, Židek L, Barda-Saad M, Chill JH (2013) NMR determines transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein. Biophys J 105:481–493

Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. Chem Rev 114:6561–6588

Heller GT, Aprile FA, Vendruscolo M (2017) Methods of probing the interactions between small molecules and disordered proteins. Cell Mol Life Sci 74:3225–3243

Hiller S, Wider G (2012) Automated projection spectroscopy and its applications. Top Curr Chem 316:21–47

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881

Hiller S, Wider G, Wüthrich K (2008) APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR 42:179–195

Hoch JC, Stern AS (2001) Nuclear magnetic resonance of biological macromolecules. Academic Press, Cambridge, pp 159–178

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in *Escherichia coli*. Protein Expr Purif 42:173–177

Joshi P, Chia S, Habchi J, Knowles TPJ, Dobson CM, Vendruscolo M (2016) A fragment-based method of creating small-molecule libraries to target the aggregation of intrinsically disordered proteins. ACS Comb Sci 18:144–153

Kadkhodaie M, Rivas O, Tan M, Mohebbi A, Shaka AJ (1991) Broadband homonuclear cross polarization using flip-flop spectroscopy. J Magn Reson 91:437–443

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc 125:1385–1393

Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes: a leap in NMR technology. Prog NMR Spectrosc 46:131–155

Kupce E, Freeman R (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. J Am Chem Soc 126:6429–6440

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUPMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. Eur J Biochem 256:1–15

Motackova V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Židek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. J Biomol NMR 48:169–177

Narayanan RL, Duerr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Židek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Nováček J, Haba NY, Chill JH, Židek L, Sklenář V (2012) 4D non-uniformly sampled HCBCACON and $^1J_{NC\alpha}$-selective HCB-CANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating ^1H chemical shift determination into ^{13}C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27:165–173

Piai A, Hošek T, Gonnelli L, Zawadzka-Kazimierczuk A, Koźmiński W, Brutscher B, Bermel W, Pierattelli R, Felli IC (2014) "CON–CON" assignment strategy for highly flexible intrinsically disordered proteins. J Biomol NMR 60:209–218

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progr NMR Spectrosc 34:93–158

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Tóth G, Gardai SJ, Zago W, Bertoncini CW, Cremades N, Roy SL, Tambe MA, Rochet JC, Galvagnion C, Skibinski G, Finkbeiner S, Bova M, Regnstrom K, Chiou SS, Johnston J, Callaway K, Anderson JP, Jobling MF, Buell AK, Yednock TA, Knowles TP, Vendruscolo M, Christodoulou J, Dobson CM, Schenk D, McConlogue L (2014) Targeting the intrinsically disordered structural ensemble of α-synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease. PLoS ONE 9:e87133

Uversky V, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246

Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. Chem Rev 114:6844–6879

van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. Chem Rev 114:6589–6631

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18–29

Yao X, Stefan B, Zweckstetter M (2014) A six-dimensional alpha proton detection-based APSY experiment for backbone assignment of intrinsically disordered proteins. J Biomol NMR 60:231–240

Zawadzka-Kazimierczuk A, Kazimierczuk K, Koźmiński W (2010) A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins. J Magn Reson 202:109–116

Zawadzka-Kazimierczuk A, Koźmiński W, Sanderová H, Krásný L (2012) High dimensional and high resolution pulse seqeunces for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52:329–337

¹³C APSY-NMR for sequential assignment of intrinsically disordered proteins

Maria Grazia Murrali[1], Marco Schiavina[1], Valerio Sainati[1], Wolfgang Bermel[2], Roberta Pierattelli[1,3,✉] and Isabella C. Felli[1,3,✉]
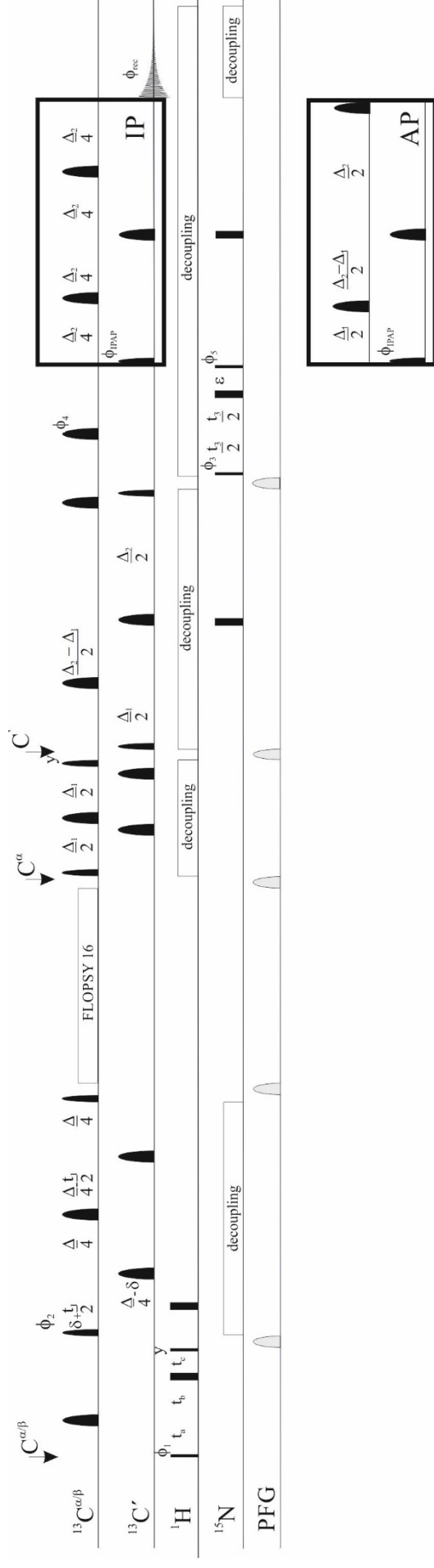
Supporting material

Figure S1.

Figure S1. Pulse scheme for the 4D CT-HCCCON. The following phase cycling was employed: ϕ_1 = x, -x ; ϕ_2 = 2(y), 2(-y); ϕ_3 = 4(x), 4(-x) ; ϕ_4 = 16(x), 16 (-x); ϕ_5= 8(x), 8(x), ϕ_{rec}=x, 2(-x), x, (-x), 2(x), (-x), 2(x), x, 2(-x), x The quadrature detection was achieved through the States-TPPI approach by incrementing the phase of the π/2 pulse prior to the evolution period; C' homonuclear decoupling was achieved using IPAP scheme. The length of the delays were: δ= 0.475 ms, Δ= 56 ms, Δ_1= 9 ms, Δ_2= 25ms – p_{12} , ε= $t_3(0)$ + p_8. t_a, t_b and t_c were used to achieve the semi-constant time mode for the ^{1}H indirect dimension. The ^{1}H carrier was placed in the middle of the ^{1}H aliphatic region (2.5 ppm) during the evolution block while it was set at 4.7 ppm in the remaining parts of the pulse sequence.

Figure S2
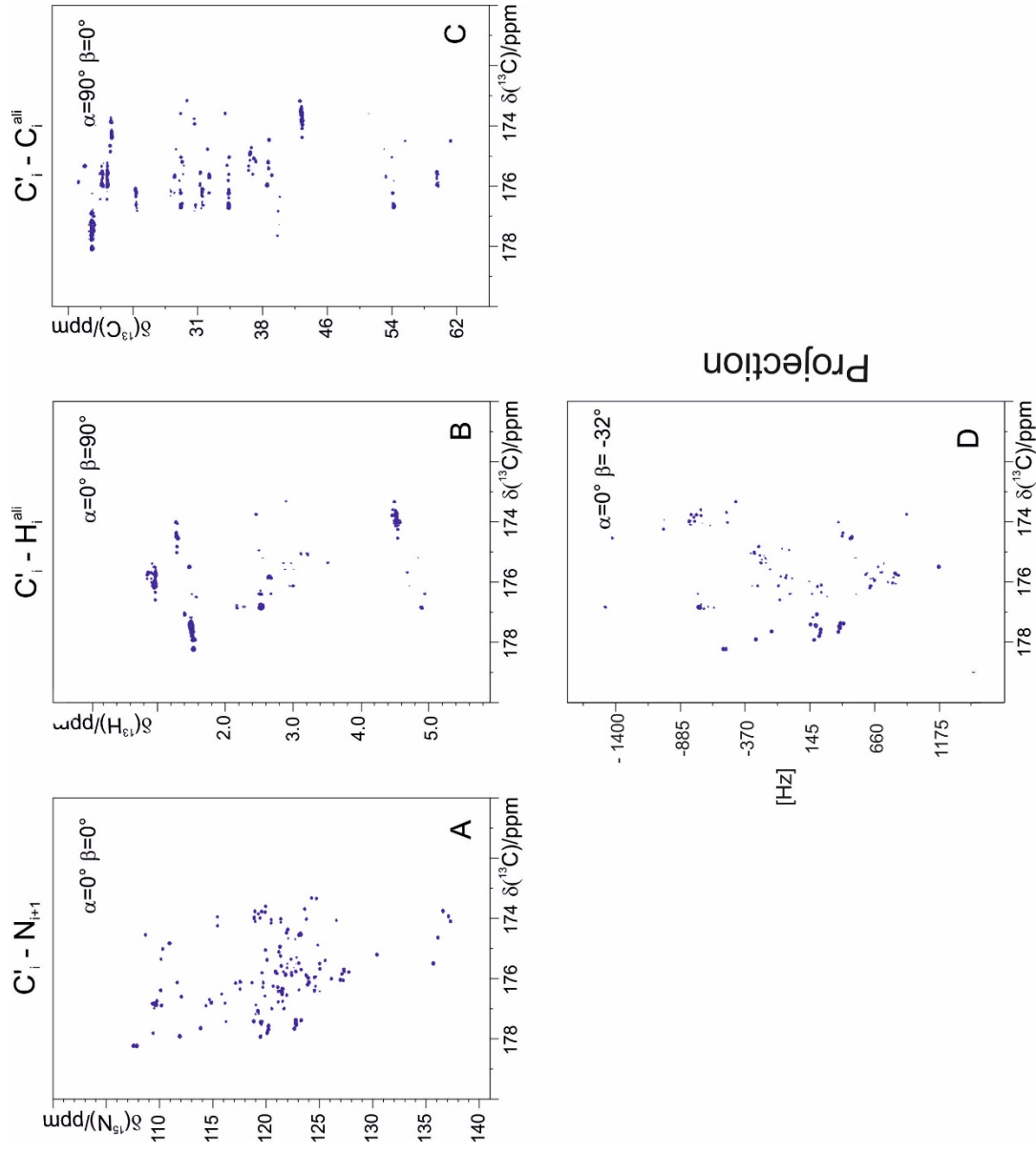
Figure S2. Examples of 2D projections acquired with the 4D CT-HCCCON experiment to show the quality of the data. Three orthogonal projections (A – C) and a non-orthogonal one (D) are shown. The spectra were recorded at 700 MHz on α-synuclein at 315 K. The three orthogonal projections correspond to C'(i)-N(i+1), C'(i)-Hali(i), C'(i)-Cali(i). The non-orthogonal projection has, as indirect dimension, the combination of $\omega_1 \sin(-32°) + \omega_3 \cos(-32°)$.

2.3 New time-saving strategies: the use of Multiple Receivers

Disordered proteins are very often affected by sample instability due to aggregation or proteolytic degradation. Even exploiting smart strategies during protein purification to get rid of proteases, limited lifetime of samples should be taken into count during the implementation of an NMR-based research project. In the previous chapters several approaches to reduce experimental time have been described.[78,85] The introduction of Multiple Receivers (MR) as standard feature within the Bruker new NMR consoles AVANCE NEO provide a novel and easily accessible way to combine multiple experiments in one.

During recent years, several applications of multiple receivers for biomolecular NMR experiments have been developed for either solution and solid state NMR.[86–89] There are two main schemes to utilize multiple receivers: the parallel scheme, in which several FIDs from different nuclear species are recorded simultaneously, and the sequential scheme, in which the different FIDs are recorded sequentially.[88,90,91]

The sequential scheme is the one first used, as it can be implemented by jumping the receiver frequency between the nuclei of interest without using MR. On the other hand, the introduction of MR makes the use of this sequential scheme more efficient and straightforward, stimulating the development and the application of this strategy.

As discussed at the beginning of this chapter, ^{13}C direct detection, and the exploitation of C'-based experiments in particular, has proven to be key to investigate IDPs. Despite the advantages, however, ^{13}C detected experiments based on C'-N correlation have also drawbacks such as the long relaxation delay required by C' magnetization to relax back to equilibrium. By using MR it is possible to profitably use the long relaxation delay of the CON experiment to acquire additional experiments. Provided that different nuclear spins are exploited and that the additional experiments do not interfere with the 2D CON itself, it is possible to combine 2D CON experiment with ^{1}H-detected experiments, such as an HN HSQC, or with ^{15}N-detected experiments, such as 2D NH-BEST-TROSY and 2D (H)CAN. Interestingly, considering the short relaxation delay needed to restore the proton magnetization in both the two ^{15}N-detected

experiments, it is possible to acquire these experiments more than once during the CON relaxation delay, overcoming the low sensitivity of ^{15}N nuclei. Technical compromises have to be introduced in some of these experiments, in particular to avoid composite pulse decoupling of carbon nuclear spins during the acquisition of the ^{15}N FID, to preserve the C' polarization for the CON experiment. Actually, these modifications slightly reduce the ^{13}C' sensitivity, but this combination is still very profitable considering that the ^{13}C is the most sensitive between the two. The total experimental time to acquire both the 2D CON and, e.g. a ^{15}N detected 2D (H)CAN, with sufficient signal to noise ratio is anyway widely reduced with respect of running the two experiments independently.

The use of MR with sequential experiments based on 2D CON as principal experiment is a very promising strategy for the study of IDPs with limited lifetime, including IDPs inside whole cells and cell lysates, or for monitoring chemical reactions at atomic level with high resolution.

2.3.1 Article: Taking simultaneous snapshots of intrinsically disordered proteins in action

Article

Biophysical Society

Taking Simultaneous Snapshots of Intrinsically Disordered Proteins in Action

Marco Schiavina,[1] Maria Grazia Murrali,[1] Letizia Pontoriero,[1] Valerio Sainati,[1] Rainer Kümmerle,[2] Wolfgang Bermel,[3] Roberta Pierattelli,[1,*] and Isabella C. Felli[1,*]

[1]Magnetic Resonance Center and Department of Chemistry "Ugo Schiff," University of Florence, Sesto Fiorentino, Florence, Italy; [2]Bruker BioSpin AG, Fällanden, Switzerland; and [3]Bruker BioSpin GmbH, Rheinstetten, Germany

ABSTRACT   Intrinsically disordered proteins (IDPs) as well as intrinsically disordered regions (IDRs) of complex protein machineries have recently been recognized as key players in many cellular functions. NMR represents a unique tool to access atomic resolution structural and dynamic information on highly flexible IDPs/IDRs. Improvements in instrumental sensitivity made heteronuclear direct detection possible for biomolecular NMR applications. The CON experiment has become one of the most useful NMR experiments to get a snapshot of an IDP/IDR in conditions approaching physiological ones. The availability of NMR spectrometers equipped with multiple receivers now enables the acquisition of several experiments simultaneously instead of one after the other. Here, we propose several variants of the CON experiment in which, during the recovery delay, a second two-dimensional experiment is acquired, either based on ^{1}H detection (CON//HN) or on ^{15}N detection (CON//btNH, CON//(H)CAN). The possibility to collect simultaneous snapshots of an IDP/IDR through different two-dimensional spectra provides a novel tool to follow chemical reactions, such as the occurrence of posttranslational modifications, as well as to study samples of limited lifetime such as cell lysates or whole cells.

SIGNIFICANCE   The exploitation of multiple receivers available in most of the newly designed NMR spectrometers enables the simultaneous collection of several snapshots of the very same event from different atomic perspectives.

INTRODUCTION

Proteins or protein regions that lack a well-defined three-dimensional (3D) structure and that are characterized by high flexibility and disorder are widespread in living organisms and essential for protein function (1–3). Their contributions to biological processes are highly complementary to those typical of folded proteins. Indeed, intrinsically disordered proteins (IDPs) or intrinsically disordered protein regions (IDRs) are often involved in key regulatory processes for which the adaptability of the protein structure and dynamics represents a clear functional advantage (4). Their important functional role also becomes evident from the strong link that has been found between misfunction in IDPs/IDRs and many diseases such as cancer and neurodegenerative diseases. The atomic-level characterization of IDPs/IDRs has thus become a topic of central importance

also in light of the development of new drugs capable of interfering with them, a completely novel area for which the traditional approach of drug design based on the identification of well-defined binding pockets in folded proteins is obviously bound to fail (5–8).

NMR plays a central role in this area of research being the only method able to provide atomic resolution information on the structural and dynamic properties of highly flexible macromolecules (9–13). NMR observables are, however, strongly influenced by the peculiar properties of IDPs, and thus the NMR approaches should be optimized to overcome critical points (14). For example, it is well known that [1]H-detected experiments, usually the first choice on the grounds of instrumental sensitivity, do have some drawbacks for the study of IDPs, in particular when approaching physiological pH and temperature conditions. In fact, the largely solvent-exposed backbones typical of highly flexible IDPs leave amide protons accessible to the solvent and are responsible for efficient chemical exchange processes that may lead to extreme broadening of H^N resonances, reducing the sensitivity of [1]H-detected experiments. In addition, ^{1}H nuclear

spins are characterized by a moderate chemical shift dispersion. These are two features that have stimulated the development of exclusively heteronuclear NMR experiments based on ^{13}C direct detection for the study of IDPs. The two-dimensional (2D) CON experiment (15,16) is now widely used to characterize highly flexible IDPs, thanks to the excellent chemical shift dispersion of the crosspeaks observed in this experiment and the possibility to reveal atomic resolution information on IDPs also in experimental conditions in which H^N resonances are not observable (17–28). The wide use of exclusively heteronuclear NMR experiments for protein investigations has been stimulated by improvements in instrumental sensitivity (29). Technological advances provided a jump of a factor of more than 10 in ^{13}C instrumental sensitivity, and lately, similar efforts have even enabled direct detection of ^{15}N (30–33), a nucleus characterized by a lower gyromagnetic ratio with respect to ^{13}C and thus by a very low intrinsic sensitivity.

In parallel to significant increase in instrumental sensitivity, new technologies have also become easily accessible such as the possibility to exploit NMR spectrometers with multiple receivers, enabling the design of multinuclear pulse sequences for the acquisition of more than one free induction decay (FID) for each repetition. Several different strategies exploiting multiple receivers have been proposed in the literature (34–39) to acquire more experiments in the time needed for a single one. Applications of multiple receivers for biomolecular NMR experiments were also proposed either in solution or in the solid state (36–41). One strategy consists of the acquisition of different FIDs within one main coherence transfer pathway to recover signals that would otherwise be lost or suppressed (36,37). The other strategy consists in exploiting polarization that remains unused at the end of one experiment to collect more than one spectrum in the time needed for the longest one of them (38).

The development of more sensitive instruments and the possibility to use multiple receivers stimulates the design of novel experimental strategies to investigate IDPs, and we will show some examples of the useful information that can be gathered combining the CON with other experimental schemes.

MATERIALS AND METHODS

Samples preparation

The ^{13}C, ^{15}N-labeled α-synuclein sample was prepared as previously described in the literature (42), with a final concentration of 1.0 mM in 100 mM NaCl, 50 μM ethylenediaminetetraacetic acid (EDTA), and 20 mM phosphate buffer at different pH between 6.5 and 7.4 with the addition of 2–10% D_2O for the lock. Each figure reports the details about sample conditions. The ^{13}C, ^{15}N-labeled α-synuclein samples used for in-cell experiments and for experiments on cell lysate were prepared as previously described (43) with an estimated protein concentration of 250 μM, using the same buffer as for the purified samples at pH 7.4.

For the phosphorylation reaction, 700 U of tyrosine kinase Fyn (Sigma Aldrich, St. Louis, MO) were added to a 200 μM sample of ^{13}C, ^{15}N-labeled

α-synuclein in 20 mM phosphate buffer, 100 mM NaCl, 50 μM EDTA, 2 mM dithiothreitol, 6 mM $MgCl_2$, and 3 mM ATP buffer at pH 7.0 for a final volume of 300 μL, with 5% D_2O added for the lock.

NMR experiments

The NMR experiments were acquired at different temperatures on a Bruker AVANCE NEO spectrometer (Billerica, MA) operating at 700.06 MHz ^{1}H, 176.05 MHz ^{13}C, and 70.97 MHz ^{15}N frequencies equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection (TXO). Standard radio frequency pulses and carrier frequencies for triple resonance experiments were used and are summarized hereafter. Q5- and Q3-shaped pulses (44) of durations of 300 and 231 μs, respectively, were used for ^{13}C band-selective $\pi/2$ and π flip angle pulses except for the π pulses that should be band selective on the C^α region (Q3, 1200 μs) and for the adiabatic π pulse to invert both C′ and C^α (smoothed chirp 500 μs, 20% smoothing, 80 kHz sweep width, 11.3 kHz radio frequency field strength) (45). An Eburp2 shape of duration of 1.768 ms and a Reburp shape of duration of 2.076 ms were used, respectively, for ^{1}H band-selective $\pi/2$ and π flip angle pulses. In the CON//HN experiment, solvent suppression was achieved through the 3:9:19 pulse scheme (46).

The ^{13}C band-selective pulses on C^α and C′ were applied at the center of each region, respectively, and the ^{1}H-band-selective pulses were applied at the center of the H^N region at 8.13 ppm. Decoupling of $^{13}C^\alpha$ and ^{15}N was achieved with p5m4sp180 (Q3, 900 μs) and Waltz65 (100 μs) decoupling sequences, respectively (44,47). All gradients employed had a smoothed square shape.

All the spectra were acquired, processed, and analyzed by using Bruker TopSpin 4.0.1 software. Calibration of the spectra was achieved using 4,4-dimethyl-4-silapentane-1-sulfonic acid as a standard for ^{1}H and ^{13}C; ^{15}N shifts were calibrated indirectly (48).

The CON//HN was acquired with a CON interscan delay of 1.7 s; the HN was acquired within this delay. For each increment of the CON experiment, the in-phase (IP) and antiphase (AP) components were acquired and properly combined to achieve IPAP virtual decoupling (49) as described in Fig. S1. The CON spectrum was acquired with two scans, with sweep widths of 5555 Hz (^{13}C) × 2500 Hz (^{15}N) and 1024 × 512 real points in the two dimensions, respectively. The HN spectrum was acquired with 4 scans (2 scans as used for the CON × 2 additional scans because no IPAP decoupling is necessary for the HN), with sweep widths of 10,869 Hz (^{1}H) × 2000 Hz (^{15}N) and 1536 × 512 real points in the two dimensions, respectively. The total duration of the experiment was 1 h and 7 min. For comparison purposes, the two independent experiments with the same parameters were also acquired.

Similar parameters were used also for CON//HN experiments acquired for cell lysates and for in-cell samples as well as to follow the phosphorylation reaction, with the main difference being the use of four scans in the latter case because of lower sample concentration.

The CON//btNH was acquired with a CON interscan delay of 2.5 s; during this time, the btNH was repeated four times (recovery delay of 400 ms). For each increment of each experiment, the IP and AP components were acquired and properly combined to achieve IPAP virtual decoupling (49) as described in Fig. S2. The CON spectrum was acquired with 4 scans, with sweep widths of 5263 Hz (^{13}C′) × 2403 Hz (^{15}N) and 1024 × 1024 real points in the 2 dimensions, respectively. The btNH spectrum was acquired with 128 scans (4 scans as used for the CON × 4 repetitions of the btNH experiment × 8 because of the different number of increments needed in the indirect dimension), with sweep widths of 5263 Hz (^{15}N) × 2403 Hz (^{1}H) and 1024 × 128 real points in the two dimensions, respectively. The total duration of the experiment was 6 h and 19 min. For comparison purposes, the two independent experiments with the same parameters were also acquired.

The CON//(H)CAN was acquired with a CON interscan delay of 2.3 s; during this time, the (H)CAN was repeated three times (recovery delay of 700 ms). For each increment of each experiment, the IP and AP

components were acquired and properly combined to achieve IPAP virtual decoupling (49) as described in Fig. S3. The CON spectrum was acquired with 8 scans, with sweep widths of 5555 (^{13}C′) × 2554 Hz (^{15}N) and 1024 × 2048 real points in the two dimensions, respectively. The (H) CAN spectrum was acquired with 192 scans (8 scans as used for the CON × 3 repetitions of the (H)CAN experiment × 8 because of the different number of increments needed in the indirect dimension), with sweep widths of 5555 Hz (^{15}N) × 4000 Hz (^{13}C) and 1024 × 256 real points in the 2 dimensions, respectively. The total duration of the experiment was 60 h and 10 min. For comparison purposes, the two independent experiments with the same parameters were also acquired.

RESULTS AND DISCUSSION

The 2D CON has by now become one of the key experiments to characterize IDPs. Many experimental variants of the basic CON pulse sequence have been developed (43,50). Important features include the nuclear spins used as starting polarization source as well as the approach employed for homonuclear decoupling. When working with sample conditions approaching the "physiological" ones, like in the presence of high ionic strength, pH above 7, and relatively high temperature, the most preferred scheme is the simplest one, which starts on ^{13}C and actively exploits heteronuclear spins only. However, it requires a relaxation delay of ∼2 s to allow the C′ magnetization to relax back to equilibrium. Can we profitably use this delay to collect additional experiments?

The CON//HN experiment

The first two spectra generally collected on an IDP are the 2D HN and 2D CON. The experimental variant designed to acquire them simultaneously is here referred to as the CON//HN, and the spectra obtained on α-synuclein are shown in Fig. 1. The pulse sequence is reported in Fig. S1 and sketched in Fig. 1 A. In brief, a very simple logic was followed in the design of the pulse sequence. The CON experiment essentially exploits only two backbone nuclear spins (C′ and N). During the relaxation delay, an additional experiment can be acquired provided that different nuclear spins are perturbed and that it does not interfere with the CON itself. The ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) exploits the amide proton as a starting polarization source as well as for acquisition of the FID; ^{15}N chemical shifts are sampled in the indirect dimension, just like in the CON experiment, and this does not cause any interference. Thus, it can be profitably acquired during the 2-s recycle delay of the CON. The only compromise to combine the two experiments derives from the need to decouple ^{15}N from ^{1}H and ^{13}C during the chemical shift evolution period. For the CON experiment, it is very important to use composite pulse decoupling of ^{1}H throughout the whole pulse sequence to minimize possible influence of exchange processes of amide protons with the solvent protons, which reduce signal intensities in particular when approach-
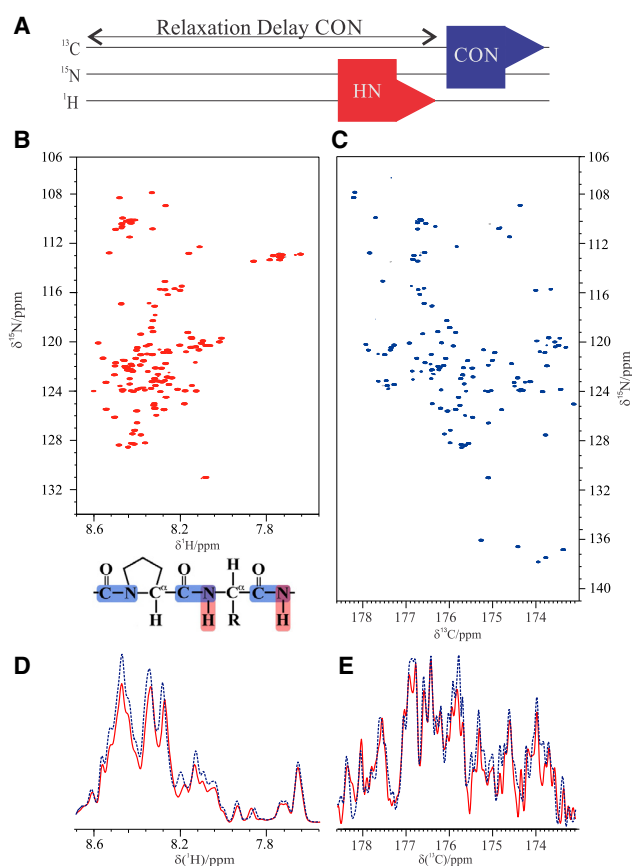

FIGURE 1 (A) Scheme of the CON//HN experiment. (B) HN and (C) CON 2D spectra acquired through the CON//HN experiment on ^{13}C, ^{15}N-labeled α-synuclein 1 mM at 285 K in 100 mM NaCl, 50 μM EDTA, and 20 mM phosphate buffer (pH 6.5) are shown. (D) A projection of the HN and the (E) CON spectra acquired with the CON//HN experiment (red solid traces) and with the corresponding experiments acquired independently (blue dotted traces) is shown. To see this figure in color, go online.

ing physiological conditions. This does not constitute a problem because the ^{1}H-^{15}N HSQC pulse sequence precedes the CON one. Decoupling of ^{15}N from ^{13}C during the ^{15}N evolution period of the ^{1}H-^{15}N HSQC could be achieved by a combination of two 180° pulses in appropriate positions to restore C′ magnetization to equilibrium as a starting polarization source for the CON experiment. Thanks to this solution, the two spectra can be acquired simultaneously in the optimal conditions required for each of them as one can see by inspecting the quality of the spectra reported in Fig. 1. These were acquired on α-synuclein at 285 K and pH 6.5, which are conditions useful to evaluate whether all the expected resonances could be observed through the CON//HN experiment. Indeed, more than 99% of the backbone correlations could be detected in both experiments (133 out of 134 in the HN and 138 out of 139 in the CON, respectively).

Comparison of the intensities of the spectra collected through the CON//HN experiment with those acquired with exactly the same parameters but in an independent fashion

shows that reduction in signal to noise ratio (S/N) resulting from the combination of the two experiments is minimal for the CON (5%) and more significant for the HN (15%) which, however, has an intrinsically higher sensitivity and provides the same number of peaks as the independent spectrum (Fig. 1, *D* and *E*).

When going to more demanding samples of limited lifetime, the advantages of recording the CON//HN are even more striking. The spectra acquired on α-synuclein in cell are shown in Fig. 2 *C* and are compared with those acquired on α-synuclein on cell lysates (Fig. 2 *B*) as well as on purified α-synuclein (Fig. 2 *A*) in conditions approaching physiological ones (310 K; pH 7.4). First of all, it is interesting to observe the spectra acquired on purified α-synuclein in these conditions and compare them with the analogous one acquired at lower pH and temperature (Fig. 1). The CON results are particularly useful to access complete information (138 detected backbone correlations; 99.3% of the expected ones); the quality of the HSQC is instead reduced because of increased solvent exchange broadening at higher pH and temperature (108 detected backbone correlations, 80.6% of the expected ones). Still, the two spectra provide interesting complementary information, including qualitative information about exchange processes.

The same spectra acquired on cell lysate show that most of the signals detected on the purified sample can be observed also in cell lysates. The possibility to acquire the two spectra simultaneously is very important, in particular for cell lysates, which have limited lifetimes. It is worth noting that the linewidths of the observed crosspeaks in the two spectra are very similar to those observed for purified α-synuclein, a nontrivial observation considering the complexity of cell lysate composition. When moving to in-cell experiments, the two spectra still show that a vast majority of the expected signals can be detected, providing atomic resolution information to characterize IDPs inside whole cells. However, the increase in linewidth when going to in-cell spectra is much more pronounced for the ^{1}H signals with respect to ^{13}C signals. Simultaneous acquisition of the two spectra is important to confirm that it is a real property of in-cell spectra and that it does not depend on possible changes in the experimental conditions or in-cell sample quality. Therefore, the CON//HN provides at the same time information derived from the exclusively heteronuclear spectra, characterized by narrow linewidths, as well as from proton-detected spectra, characterized by higher sensitivity but more influenced by line broadening deriving from both the inhomogeneous environment as well as from exchange processes.

The CON//^{15}N-detected experiments

^{15}N detection has interesting features for the study of IDPs because of the narrow linewidths and large dispersion of chemical shifts, provided one is in the appropriate sensitivity regime, which nowadays still requires highly
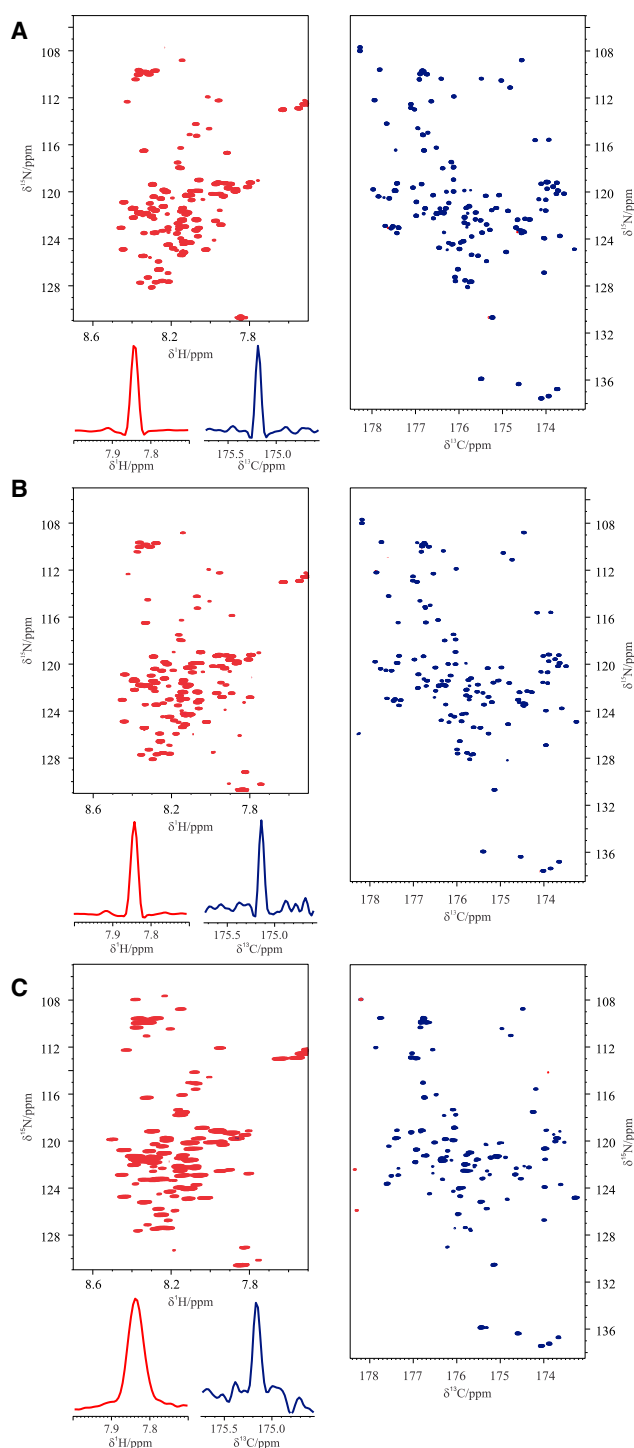

FIGURE 2 Comparison of the 2D spectra (HN *left*; CON *right*) acquired through the CON//HN experiment on ^{13}C, ^{15}N-labeled α-synuclein at 310 K in different experimental conditions: (*A*) purified sample in 100 mM NaCl, 50 μM EDTA, and 20 mM phosphate buffer at pH 7.4 is shown; (*B*) in *Escherichia coli* cells lysate resuspended in the same buffer as in (*A*) and (*C*) in-cell. The traces of a representative signal extracted from the HN (*red*; *left*) and the CON (*blue*; *right*) spectra are also reported. To see this figure in color, go online.

concentrated samples ([30]–[33]). The CON experiment was first combined with the ^{15}N-detected experiment to collect 2D HN spectra (CON//btNH) ([51]). The BEST-TROSY variant ([52]), which uses band-selective pulses for the amide protons, was selected to exploit the very efficient longitudinal relaxation enhancement of amide protons in IDPs ([53]–[55]) as well as to avoid ^{1}H decoupling during acquisition of the ^{15}N FID, two features that enable the acquisition of 3–4 ^{15}N FIDs during the CON relaxation delay to overcome the very low sensitivity of ^{15}N. Decoupling of ^{1}H during the ^{15}N evolution delay of the CON was achieved by pairs of 180° H^N band-selective pulses to restore the starting polarization source for the BEST-TROSY experiments. In addition, there is no proton decoupling during the acquisition of the CON. These two features reintroduce some dependence on solvent exchange processes as well as C′-H^N couplings in the direct acquisition dimension. The major modification included in the BEST-TROSY part of the experiment was to avoid composite pulse decoupling of carbonyl carbon nuclear spins during the acquisition of the ^{15}N FID to preserve C′ polarization for the following CON experiment. To this end, virtual decoupling of C′ during the acquisition of the ^{15}N FID was implemented through the IPAP approach ([56]) in combination with band-selective decoupling of the C$^\alpha$ spins. The pulse sequence is reported in Fig. S2 and schematically illustrated in Fig. 3 A. The results of the CON//btNH experiments recorded on α-synuclein at different temperatures are reported in Fig. 3, B and C. The good performance of the experiment is shown by the well-resolved spectra that can be acquired; essentially all expected correlations can be detected at 285 K and pH 6.5 (138 detected backbone correlations for the CON and 133 for the btNH; 99.3% of the expected ones for each spectrum). The modifications introduced to acquire the two spectra simultaneously slightly reduce the sensitivity of the CON spectrum, which is, however, the more sensitive of the two and thus only has a minor impact on the final outcome (Fig. 3, D and E). The ^{15}N-detected BEST-TROSY HN spectrum, although less influenced by exchange broadening of amide protons than its ^{1}H-detected counterpart, still relies on amide protons in the initial coherence transfer step as well as in the ^{1}H chemical shift evolution achieved in the indirect dimension. As expected, with increasing temperature, the quality of the CON is essentially maintained, whereas the intensity of crosspeaks in the ^{15}N-detected BEST-TROSY spectrum is reduced. Still, it can be observed that even at higher temperature than physiological one (315 K), a few additional crosspeaks can be observed in the ^{15}N-detected BEST-TROSY HN spectrum that were not identified in these conditions through ^{1}H-detected experiments (Biological Magnetic Resonance Bank [BMRB]: 27348) ([57]). In particular, in the ^{1}H-detected experiment, 93/134 peaks were visible, whereas when performing the ^{15}N-detected variant, seven additional peaks are visible (thus 100/134), mostly due to polar resi-
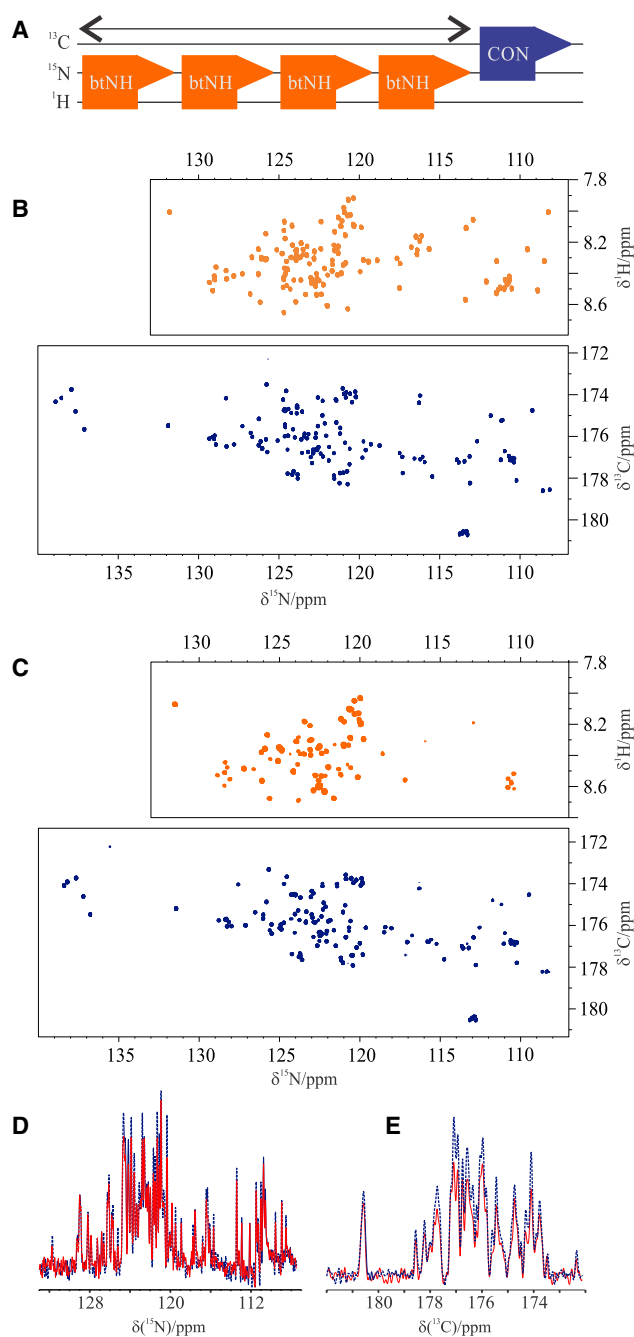

FIGURE 3   (A) Scheme of the CON//btNH experiment. The 2D spectra (btNH *top*; CON *bottom*) acquired with the CON//btNH experiment on ^{13}C,^{15}N-labeled α-synuclein 1 mM at (B) 285 K and (C) 315 K at pH 6.5 in the same buffer reported in Fig. 1 are shown. (D) and (E) report the comparison of the S/N of the projection of the spectra acquired with the CON//btNH experiment (*red solid traces*) with the analogous spectra acquired independently (*blue dotted traces*). To see this figure in color, go online.

dues (Thr-22, -64, -44, -33, Ser-9, Lys-43, Ala-76). Even if the signal intensity is lower in the ^{15}N-detected experiment because of the lower gyromagnetic ratio of ^{15}N, it is possible to collect a few more signals because the proton magnetization is transverse for a shorter amount of time,
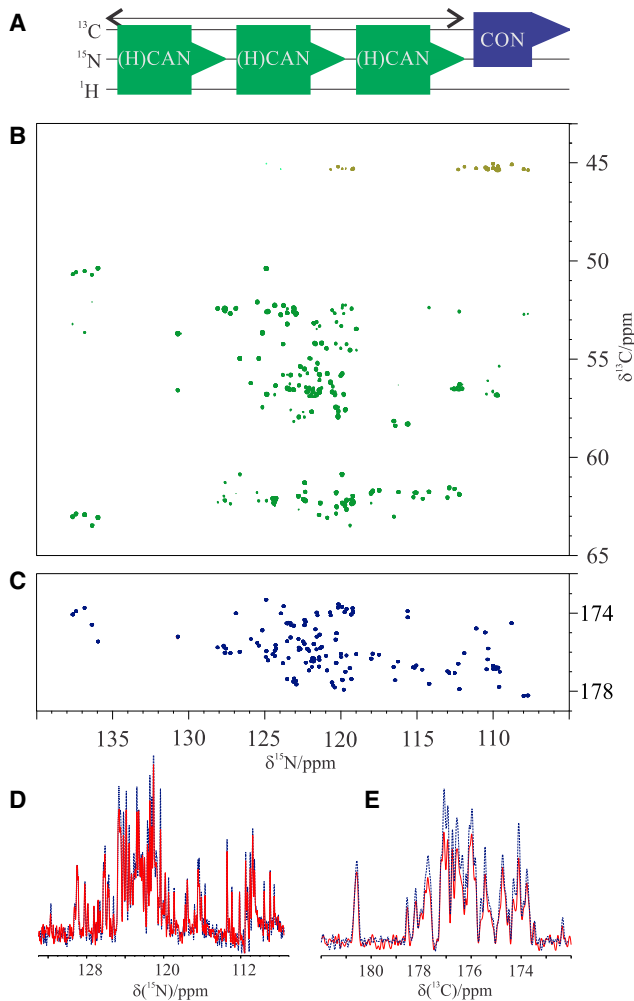
FIGURE 4 (A) Scheme of the CON//(H)CAN experiment. (B) (H)CAN and (C) CON spectra recorded on ^{13}C, ^{15}N-labeled α-synuclein at 310 K (pH 7.4) are shown. (D) and (E) report the comparison of the S/N of the projection of the spectra acquired with the CON//(H)CAN experiment (*red solid traces*) with the analogous spectra acquired independently (*blue dotted traces*). To see this figure in color, go online.

reducing losses. This shows that the method has some potential, in particular using optimized hardware (high fields and optimized probes). Another experiment based on ^{15}N detection that was combined with the CON is the (H)CAN (58). The CON//(H)CAN acquired at 310 K and pH 7.4 is shown in Fig. 4. The pulse sequence is reported in Fig. S3 and schematically illustrated in Fig. 4 *A*. As for the previous experiment, this magnetization transfer pathway starts from a proton, the H$^\alpha$, and is then transferred to the C$^\alpha$. The C$^\alpha$ dimension evolves as the indirect one. The magnetization can now be transferred to N of the same amino acid, through the ^{1}J constant, and to the previous one, through the ^{2}J constant. The two constants are similar and it is thus difficult to distinguish one pathway from the other. However, we decided to prioritize the intraresidue correlations, which are more sensitive and provide complementary information, with respect to the one available through the CON experi-

ment. When exploiting H$^\alpha$ as starting polarization source, the results are optimal at any temperature because of the nonexchangeable property of this atom. Longitudinal relaxation enhancements are expected to be modest for H$^\alpha$ in IDPs (55), and thus nonselective ^{1}H pulses can be used. More than one repetition of this experiment during the CON relaxation delay can be collected thanks to the relatively short time needed for H$^\alpha$ to recover to the steady state (hundreds of milliseconds). Technical compromises needed in this case are very similar to those discussed for the previous experiment. They are responsible for a small reduction in the sensitivity of the CON (Fig. 4, *D* and *E*), which, however, is the more sensitive between the two. Decoupling of ^{1}H and ^{13}C was achieved by pairs of 180° pulses to restore the magnetization to equilibrium for subsequent experiments. For the same reason, a virtual CO decoupling scheme is required for the acquisition of ^{15}N FIDs. To this end, we implemented the IPAP scheme, in combination with bandselective decoupling of the C$^\alpha$ spins.

In this spectrum, all the correlations of the intraresidue crosspeaks are observed as well as part of the interresidue ones (113/139; 81.0%). However, significant overlap still occurs, and only a fraction of them are resolved. It is worth noting however that with increasing magnetic fields and with tailored probes, the CON//(H)CAN becomes a promising tool for the study of IDPs. Finally, should long experimental times be necessary to compensate for a lower ^{15}N sensitivity, the CON//(H)CAN experiment can reveal weak crosspeaks present in the CON such as those deriving from the *cis-trans* proline isomerization, which has not been extensively studied for IDPs and has been found to be important in many regulatory processes.

Comparison with other methods

Here, we have chosen to opt for the most straightforward approach to combine two experiments: instead of simply waiting for the magnetization to recover before the next transient, we use the recovery delay of one particularly useful experiment, which would be acquired anyway, to collect an additional one providing complementary information for free. This method follows the idea of activating unexploited magnetization (UTOPIA) (34) but shifts the central interest to the ^{13}C-detected experiment, as schematically indicated in Fig. 1. The simplest variant of the CON experiment was selected because it represents the most robust one to ensure detection of all signals of an IDP near physiological conditions, including those that are invisible in HN spectra (residues experiencing fast solvent exchange and proline residues). This was combined either with ^{1}H-detected or with ^{15}N-detected 2D experiments.

In principle, a 3D HNCO (or a 3D HNCA), which can nowadays be collected in a very quick time thanks to projection spectroscopy (59) or nonuniform acquisition and processing strategies (60,61), could provide very similar

information. However, the results of these 3D spectra depend on the observability of amide protons. Therefore, the approach presented here is more suitable to study IDPs approaching physiological conditions (310 K; pH 7.4; ionic strength value around 0.25 M; in-cell environment) because the CON is not affected by fast solvent exchange processes.

Similar arguments hold when comparing the current approach with an elegant experimental variant of the HNCO proposed for parallel acquisition of the HN and (H^N)CON experiments (33). In this case, the two spectra (HN and (H^N)CON) are obtained by recovering magnetization from two different pathways that however both originate from H^N polarization. As a result, the two spectra are influenced by solvent exchange broadening, and the (H^N) CON does not reveal proline signals.

The relative sensitivity of the two experiments to be combined is an important issue to be considered. There is no doubt that the intrinsic sensitivity of ^{13}C is lower than that of ^1H because of the lower gyromagnetic ratio. However, thanks to instrumental improvements, the sensitivity of the CON experiments acquired on highly flexible IDPs is sufficient to collect the experiment with just a few scans per increment. Therefore, it is worth combining it with the other experiment that is generally collected on an IDP, the HN correlation spectrum. The relative sensitivity of the two experiments is inverted when combining the CON with ^{15}N-detected experiments. This difference can be partly compensated by exploiting ^1H as a starting polarization source, exploiting longitudinal relaxation enhancement (when possible) (49,51,52,55), or accumulating more ^{15}N FIDs. A leap in instrumental sensitivity exploiting dedicated probe heads will provide the necessary push to make this combination of experiments really amenable for every application.

From this analysis, it is clear that more than for time savings, which are generally moderate when considering combinations of simple 2D experiments, the simultaneous acquisition of two spectra is important for accessing a snapshot of an IDP in cases in which the sample either has limited stability in time or it changes because of chemical reactions as it happens when studying posttranslational modifications (PTMs). In these cases, to investigate the IDP of interest, in addition to the CON one can access complementary information through the multiple receiver CON variants proposed here.

A case study for PTMs

PTMs represent the most efficient way to modulate the biological activity of proteins after their biosynthesis through chemical reactions, catalyzed by enzymes. Indeed, it is possible to modulate the chemo-physical properties of a protein and thus its function by adding or removing specific chemical groups or through the cleavage of chemical bonds.

The diversity of PTMs found in proteins arises from the very different processes that they can modulate in a cell such as protein folding, signaling, molecular recognition, interactions, and degradation. In this way, the same protein can be handled for different aims, ranging from cell fate control to regulation of metabolism (62).

IDPs are frequently involved in regulatory and signaling functions, which can be modulated by PTMs (63–65). PTMs can trigger both local and long-range changes as well as intermolecular interactions. It has been supposed that structural flexibility could also be a biological strategy to overcome the classical problem of the "one lock/one key" model. The lack of a stable 3D structure guarantees adaptability for the enzymatic site and allows high specificity of the process, maintaining, however, low affinity (64,66).

Among a variety of possible modifications, one of the most common is phosphorylation. It was estimated that at least 2% of the human genome is constituted by kinases and more than 30% of the eukaryotic proteins are involved in phosphorylation, with 700,000 sites that can potentially act as substrates (67). Investigating this mechanism is fundamental to understand functions of these important enzymes. PTMs usually involve only a few residues, and NMR provides a unique tool to study these processes, giving access to residue-specific atomic resolution information.

It is well known that the role of some IDPs or IDRs and their link to the onset of several diseases is strictly dependent on PTMs, as phosphorylation (64–66,68). A clear example is the supposed key role of phosphorylation of specific α-synuclein residues in the modulation of Lewy bodies formation. In this widely studied process, serine 129 is selectively and extensively phosphorylated, promoting fibrillation in vitro (69–71). It has emerged that this process occurs together with phosphorylation of tyrosine 125, a priming event in the efficient modification of serine 129 by CK1 kinase, both in vitro under physiological conditions and in vivo (72–74).

To provide an example of the utility of the CON//HN experiment, we show here the phosphorylation reaction of tyrosine residues by Fyn near physiological conditions. Fyn is a nonreceptor tyrosine kinase of the Src family that has been shown to react with all the four tyrosine residues present in α-synuclein (Y39, Y125, Y133, Y136) with high-percentage levels but preferentially targeting Y125. The spectra acquired with the CON//HN experiment during the phosphorylation reaction are shown in Fig. 5. It can be observed that the changes in HN crosspeaks agree with previous observations (73) and that additional details are monitored through the CON spectrum acquired simultaneously (Fig. 5). In particular, the CON spectrum, which exploits the ^{13}C in the direct dimension, has a higher resolution with respect to the HN spectrum. Thus, we can also monitor the shift of those peaks, which, in the HN spectrum, are
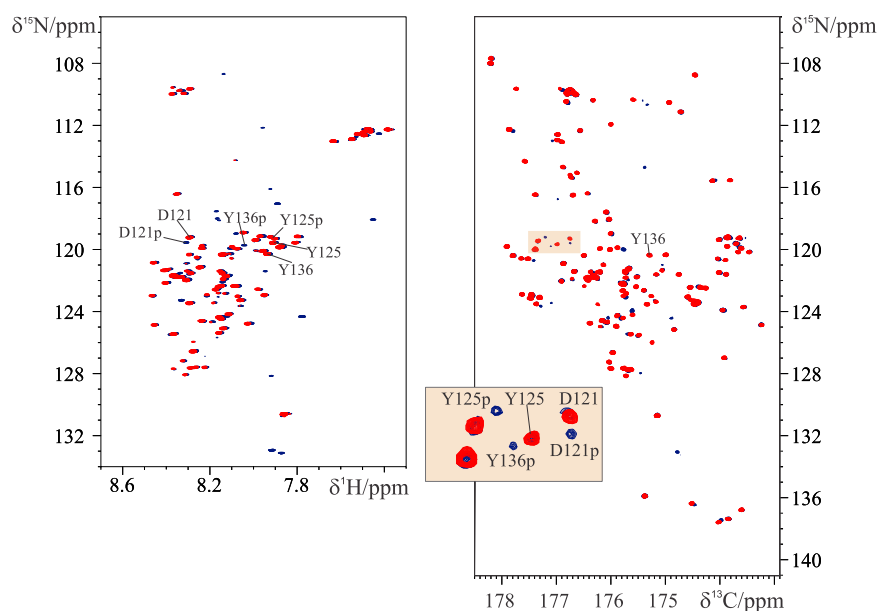
FIGURE 5 The 2D spectra acquired with the CON//HN experiment to monitor phosphorylation of ^{13}C, ^{15}N-labeled α-synuclein 200 μM with Fyn tyrosine kinase as a function of time ($t = 0$ h, *red spectra*; $t = 50$ h *blue spectra*). The HN spectra are reported on the left and the CON on the right. The highlighted region of the CON spectra is enlarged to show, as an example, a subset of α-synuclein signals that are influenced by phosphorylation. The reaction was performed at 303 K in 20 mM phosphate buffer, 100 mM NaCl, 50 μM EDTA, 2 mM dithiothreitol, 6 mM MgCl$_2$, and 3 mM ATP (pH 7.0). To see this figure in color, go online.

found in crowded regions such as those of Tyr125, which is well resolved in the CON (Fig. 5).

A final comment is necessary on the specific conditions needed to follow PTMs. Most enzymes indeed fulfill their roles at physiological conditions. Therefore, the CON//HN experiment offers a particularly useful tool because in addition to the HN, often used to study PTMs, it exploits the CON, which is not influenced by exchange broadening because of the chemical properties of the nuclei participating in the magnetization transfer pathway. The fairly low substrate concentration generally required for these reactions is still sufficient to acquire good quality ^{13}C-detected spectra, demonstrating the general applicability of this approach.

CONCLUSION

Research in the area of IDPs is continuously expanding. Partly or completely disordered proteins of increasing size and complexity and with high biomedical relevance are being discovered at fast pace, adding a novel layer to structural biology. In this frame, NMR spectroscopy has a central role, enabling the investigation of systems of increasing complexity in an environment that resembles more and more the physiological one. Heteronuclear direct detection has contributed to establishing the importance of NMR in this area of research. The possibility to collect two spectra simultaneously is very attractive to study IDPs. Several experimental variants are presented here in which during the recovery delay needed to acquire the CON, an additional 2D spectrum is collected either based on ^{1}H detection (CON//HN) or on ^{15}N detection (CON//btNH, CON//(H)CAN). This, to our knowledge, novel approach is particularly useful to study samples of limited lifetime, such as IDPs inside whole cells or cell lysates, as well as to follow chemical reactions such as PTMs of IDPs in conditions

approaching physiological ones (pH, temperature, ionic strength, inhomogeneous in-cell environment). The availability of cryogenically cooled probeheads optimized for heteronuclear direct detection is important to alleviate the problem of the intrinsic lower sensitivity of heteronuclei. The usefulness of these experiments is bound to increase with increasing magnetic fields to take maximal advantage of the narrow linewidths of C′ and N of highly flexible IDPs, enabling the study of IDPs of increasing complexity.

REFERENCES

1. Habchi, J., P. Tompa, …, V. N. Uversky. 2014. Introducing protein intrinsic disorder. *Chem. Rev.* 114:6561–6588.

2. Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16:18–29.

3. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.

4. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 37:215–246.

5. Ambadipudi, S., and M. Zweckstetter. 2016. Targeting intrinsically disordered proteins in rational drug discovery. *Expert Opin. Drug Discov.* 11:65–77.

6. Joshi, P., S. Chia, …, M. Vendruscolo. 2016. A fragment-based method of creating small-molecule libraries to target the aggregation of intrinsically disordered proteins. *ACS Comb. Sci.* 18:144–153.

7. Heller, G. T., F. A. Aprile, and M. Vendruscolo. 2017. Methods of probing the interactions between small molecules and disordered proteins. *Cell. Mol. Life Sci.* 74:3225–3243.

8. Tsafou, K., P. B. Tiwari, …, J. A. Toretsky. 2018. Targeting intrinsically disordered transcription factors: changing the paradigm. *J. Mol. Biol.* 430:2321–2341.

9. Eliezer, D. 2009. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19:23–30.

10. I. C. Felli and R. Pierattelli, eds 2015. Intrinsically Disordered Proteins Studied by NMR Spectroscopy Springer, Basel, Switzerland.

11. Gil, S., T. Hošek, …, I. C. Felli. 2013. NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. *Angew. Chem. Int. Ed. Engl.* 52:11808–11812.

12. Dyson, H. J., and P. E. Wright. 2001. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* 339:258–270.

13. Konrat, R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* 241:74–85.

14. Brutscher, B., I. C. Felli, …, Z. Sólyom. 2015. NMR methods for the study of instrinsically disordered proteins structure, dynamics, and interactions: general overview and practical guidelines. *Adv. Exp. Med. Biol.* 870:49–122.

15. Bermel, W., I. Bertini, …, R. Pierattelli. 2006. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc.* 128:3918–3919.

16. Bermel, W., I. Bertini, …, R. Pierattelli. 2006. Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *J. Magn. Reson.* 178:56–64.

17. Pérez, Y., M. Gairí, …, P. Bernadó. 2009. Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *J. Mol. Biol.* 391:136–148.

18. Knoblich, K., S. Whittaker, …, U. Günther. 2009. Backbone assignment of the N-terminal polyomavirus large T antigen. *Biomol. NMR Assign.* 3:119–123.

19. Contreras-Martos, S., A. Piai, …, P. Tompa. 2017. Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci. Rep.* 7:4676.

20. Hošek, T., E. O. Calçada, …, R. Pierattelli. 2016. Structural and dynamic characterization of the molecular hub early region 1A (E1A) from human adenovirus. *Chemistry.* 22:13010–13013.

21. Motáčková, V., J. Nováček, …, V. Sklenář. 2010. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J. Biomol. NMR.* 48:169–177.

22. Haba, N. Y., R. Gross, …, J. H. Chill. 2013. NMR determines transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein. *Biophys. J.* 105:481–493.

23. Nováček, J., N. Y. Haba, …, V. Sklenář. 2012. 4D non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J. Biomol. NMR.* 53:139–148.

24. Lawrence, C. W., and S. A. Showalter. 2012. Carbon-detected ^{15}N NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. *J. Phys. Chem. Lett.* 3:1409–1413.

25. Sahu, D., M. Bastidas, and S. A. Showalter. 2014. Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal. Biochem.* 449:17–25.

26. Lopez, J., R. Schneider, …, G. Lippens. 2016. Studying intrinsically disordered proteins under true in vivo conditions by combined cross-polarization and carbonyl-detection NMR spectroscopy. *Angew. Chem. Int.Engl.* 55:7418–7422.

27. Piai, A., E. O. Calçada, …, R. Pierattelli. 2016. Just a flexible linker? The structural and dynamic properties of CBP-ID4 revealed by NMR spectroscopy. *Biophys. J.* 110:372–381.

28. Eftekharzadeh, B., A. Piai, …, X. Salvatella. 2016. Sequence context influences the structure and aggregation behavior of a PolyQ tract. *Biophys. J.* 110:2361–2366.

29. Kovacs, H., D. Moskau, and M. Spraul. 2005. Cryogenically cooled probes - a leap in NMR technology. *Prog. Nucl. Mag. Res. Sp.* 46:131–155.

30. Takeuchi, K., G. Heffron, …, G. Wagner. 2010. Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *J. Biomol. NMR.* 47:271–282.

31. Takeuchi, K., H. Arthanari, and G. Wagner. 2016. Perspective: revisiting the field dependence of TROSY sensitivity. *J. Biomol. NMR.* 66:221–225.

32. Chhabra, S., P. Fischer, …, H. Arthanari. 2018. ^{15}N detection harnesses the slow relaxation property of nitrogen: delivering enhanced resolution for intrinsically disordered proteins. *Proc. Nat. Acad. Sci. USA.* 115:1710–1719.

33. Gibbs, E. B., and R. W. Kriwacki. 2018. Direct detection of carbon and nitrogen nuclei for high-resolution analysis of intrinsically disordered proteins using NMR spectroscopy. *Methods.* 138-139:39–46.

34. Kupče, E., R. Freeman, and B. K. John. 2006. Parallel acquisition of two-dimensional NMR spectra of several nuclear species. *J. Am. Chem. Soc.* 128:9606–9607.

35. Kupče, E., and R. Freeman. 2010. Molecular structure from a single NMR sequence (fast-PANACEA). *J. Magn. Reson.* 206:147–153.

36. Chakraborty, S., S. Paul, and R. V. Hosur. 2012. Simultaneous acquisition of ^{13}C$^{\alpha}$-^{15}N and ^{1}H-^{15}N-^{15}N sequential correlations in proteins: application of dual receivers in 3D HNN. *J. Biomol. NMR.* 52:5–10.

37. Kupče, E., and L. E. Kay. 2012. Parallel acquisition of multi-dimensional spectra in protein NMR. *J. Biomol. NMR.* 54:1–7.

38. Viegas, A., T. Viennet, …, M. Etzkorn. 2016. UTOPIA NMR: activating unexploited magnetization using interleaved low-gamma detection. *J. Biomol. NMR.* 64:9–15.

39. Kupče, E. 2013. NMR with multiple receivers. *Top. Curr. Chem.* 335:71–96.

40. Sharma, K., P. K. Madhu, and K. R. Mote. 2016. A suite of pulse sequences based on multiple sequential acquisitions at one and two radiofrequency channels for solid-state magic-angle spinning NMR studies of proteins. *J. Biomol. NMR.* 65:127–141.

41. Kupče, Ē., L. E. Kay, and R. Freeman. 2010. Detecting the "afterglow" of ^{13}C NMR in proteins using multiple receivers. *J. Am. Chem. Soc.* 132:18008–18011.

42. Huang, C., G. Ren, …, C. C. Wang. 2005. A new method for purification of recombinant human α-synuclein in *Escherichia coli. Protein Expr. Purif.* 42:173–177.

43. Felli, I. C., L. Gonnelli, and R. Pierattelli. 2014. In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat. Protoc.* 9:2005–2016.

44. Emsley, L., and G. Bodenhausen. 1992. Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *J. Magn. Reson.* 97:135–148.

45. Böhlen, J. M., and G. Bodenhausen. 1993. Experimental aspects of Chirp NMR spectroscopy. *J. Magn. Reson.* 102:293–301.

46. Piotto, M., V. Saudek, and V. Sklenár. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR.* 2:661–665.

47. Freeman, R., and H. Geen. 1991. Band-selective radiofrequency pulses. *J. Magn. Reson.* 93:93–141.

48. Markley, J. L., A. Bax, …, K. Wüthrich. 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl. Chem.* 70:117–142.

49. Felli, I. C., and R. Pierattelli. 2015. Spin-state-selective methods in solution- and solid-state biomolecular ^{13}C NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 84–85:1–13.

50. Felli, I. C., and R. Pierattelli. 2014. Novel methods based on ^{13}C detection to study intrinsically disordered proteins. *J. Magn. Reson.* 241:115–125.

51. Takeuchi, K., H. Arthanari, …, G. Wagner. 2015. Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR. *J. Biomol. NMR.* 63:323–331.

52. Solyom, Z., M. Schwarten, …, B. Brutscher. 2013. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR.* 55:311–321.

53. Schanda, P., V. Forge, and B. Brutscher. 2006. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn. Reson. Chem.* 44:S177–S184.

54. Schanda, P., H. Van Melckebeke, and B. Brutscher. 2006. Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* 128:9042–9043.

55. Hošek, T., S. Gil-Caballero, …, I. C. Felli. 2015. Longitudinal relaxation properties of ^{1}H^N and ^{1}H$^\alpha$ determined by direct-detected ^{13}C NMR experiments to study intrinsically disordered proteins (IDPs). *J. Magn. Reson.* 254:19–26.

56. Bermel, W., I. Bertini, …, R. Pierattelli. 2006. ^{13}C-detected protonless NMR spectroscopy of proteins in solution. *Prog. Nucl. Mag. Res. Sp.* 48:25–45.

57. Murrali, M. G., M. Schiavina, …, I. C. Felli. 2018. ^{13}C APSY-NMR for sequential assignment of intrinsically disordered proteins. *J. Biomol. NMR.* 70:167–175.

58. Gal, M., K. A. Edmonds, …, G. Wagner. 2011. Speeding up direct ^{15}N detection: hCaN 2D NMR experiment. *J. Biomol. NMR.* 51:497–504.

59. Hiller, S., F. Fiorito, …, G. Wider. 2005. Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci. USA.* 102:10876–10881.

60. Robson, S., H. Arthanari, …, G. Wagner. 2019. Nonuniform sampling for NMR spectroscopy. *Methods Enzymol.* 614:263–291.

61. Kazimierczuk, K., J. Stanek, …, W. Koźmiński. 2010. Random sampling in multidimensional NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 57:420–434.

62. Cheng, H. C., R. Z. Qi, …, H. J. Zhu. 2011. Regulation and function of protein kinases and phosphatases. *Enzyme Res.* 2011:794089.

63. Zhou, J., S. Zhao, and A. K. Dunker. 2018. Intrinsically disordered proteins link alternative splicing and post-translational modifications to complex cell signaling and regulation. *J. Mol. Biol.* 430:2342–2359.

64. Darling, A. L., and V. N. Uversky. 2018. Intrinsic disorder and post-translational modifications: the darker side of the biological dark matter. *Front. Genet.* 9:158.

65. Bah, A., and J. D. Forman-Kay. 2016. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* 291:6696–6705.

66. Uversky, V. N. 2014. Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. *Front. Mol. Biosci.* 1:6.

67. Ubersax, J. A., and J. E. Ferrell, Jr. 2007. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* 8:530–541.

68. Fujiwara, H., M. Hasegawa, …, T. Iwatsubo. 2002. α-synuclein is phosphorylated in synucleinopathy lesions. *Nat. Cell Biol.* 4:160–164.

69. Choi, H. S., H. Liew, …, Y. H. Suh. 2012. Phosphorylation of α-synuclein is crucial in compensating for proteasomal dysfunction. *Biochem. Biophys. Res. Commun.* 424:597–603.

70. Mahul-Mellier, A. L., B. Fauvet, …, H. A. Lashuel. 2014. c-Abl phosphorylates α-synuclein and regulates its degradation: implication for α-synuclein clearance and contribution to the pathogenesis of Parkinson's disease. *Hum. Mol. Genet.* 23:2858–2879.

71. Schmid, A. W., B. Fauvet, …, H. A. Lashuel. 2013. α-synuclein post-translational modifications as potential biomarkers for Parkinson disease and other synucleinopathies. *Mol. Cell. Proteomics.* 12:3543–3558.

72. Nakamura, T., H. Yamashita, …, S. Nakamura. 2001. Activated Fyn phosphorylates α-synuclein at tyrosine residue 125. *Biochem. Biophys. Res. Commun.* 280:1085–1092.

73. Kosten, J., A. Binolfi, …, P. Selenko. 2014. Efficient modification of α-synuclein serine 129 by protein kinase CK1 requires phosphorylation of tyrosine 125 as a priming event. *ACS Chem. Neurosci.* 5:1203–1208.

74. Paleologou, K. E., A. W. Schmid, …, H. A. Lashuel. 2008. Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of α-synuclein. *J. Biol. Chem.* 283:16895–16905.

Supplemental Information

Taking Simultaneous Snapshots of Intrinsically Disordered Proteins in Action

Marco Schiavina, Maria Grazia Murrali, Letizia Pontoriero, Valerio Sainati, Rainer Kümmerle, Wolfgang Bermel, Roberta Pierattelli, and Isabella C. Felli

Supporting Material

Taking Simultaneous Snapshots of Intrinsically Disordered Proteins in Action

M. Schiavina, M. G. Murrali, L. Pontoriero, V. Sainati, R. Kümmerle, W. Bermel, R. Pierattelli, I. C. Felli

TABLE OF CONTENTS

Figure S1. CON//HN pulse sequence scheme.

Narrow and wide black rectangles represent $\pi/2$ and π non-selective pulses; narrow and wide black symbols represent $\pi/2$ and π band-selective pulses (Q5 and Q3 shapes of 300 and 231 µs respectively), hatched pulses are adiabatic inversion pulse on C' and C$^\alpha$ (smoothed Chirp 500 µs 80 kHz sweep and 20% smooting).

The following phase cycling was employed for the CON//HN: ϕ_1 = x, -x ; ϕ_2 = 2(x), 2(-x); ϕ_3 = 4(x), 4(-x), ϕ^{IP} = x ; ϕ^{AP} = -y and ϕ_{rec}^{CON}= x, -x, x, -x, -x, x, -x, x for the 2D-CON and ϕ_4 = x, −x, ϕ_5 = 2(x), 2(-x) ϕ_6 = 4(x) 4(-x) and ϕ_{rec}^{HN} = x, -x, x, -x, -x, x, -x, x for the 2D-HN. Quadrature detection in the indirect dimension was achieved through the STATES-TPPI approach incrementing phase ϕ_1 (CON) and ϕ_4 (HN).

The length of the delays was: d_1= 1.7 s; Δ= 16.6 ms; Δ_1 = 4.5 ms; Δ_2= 2.7 ms; ε = ½ duration of adiabatic pulse (250 µs); ε'= $t_1(0)$ + duration of adiabatic pulse (500 µs).

Virtual decoupling of the C'-C$^\alpha$ coupling in the 2D-CON was achieved by acquiring for each increment both the IP and AP component of the signals and combining them.
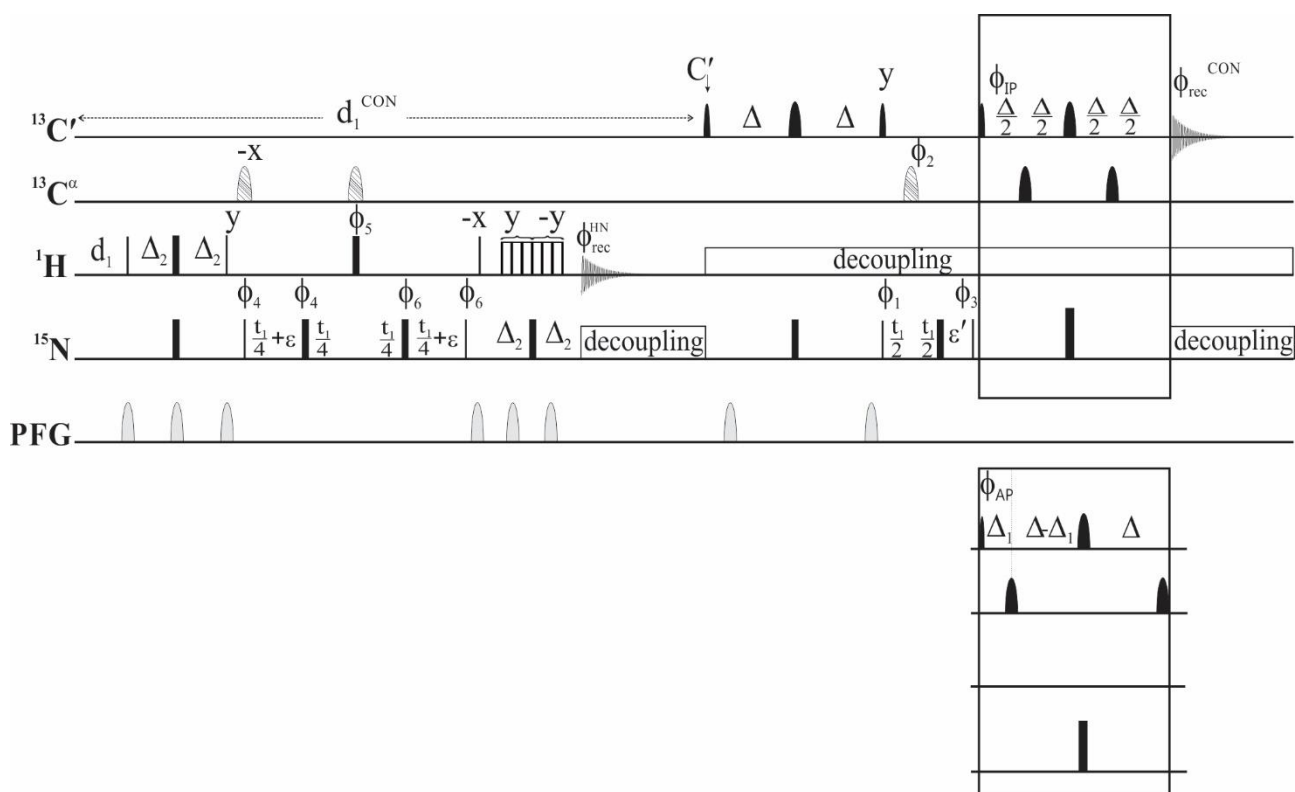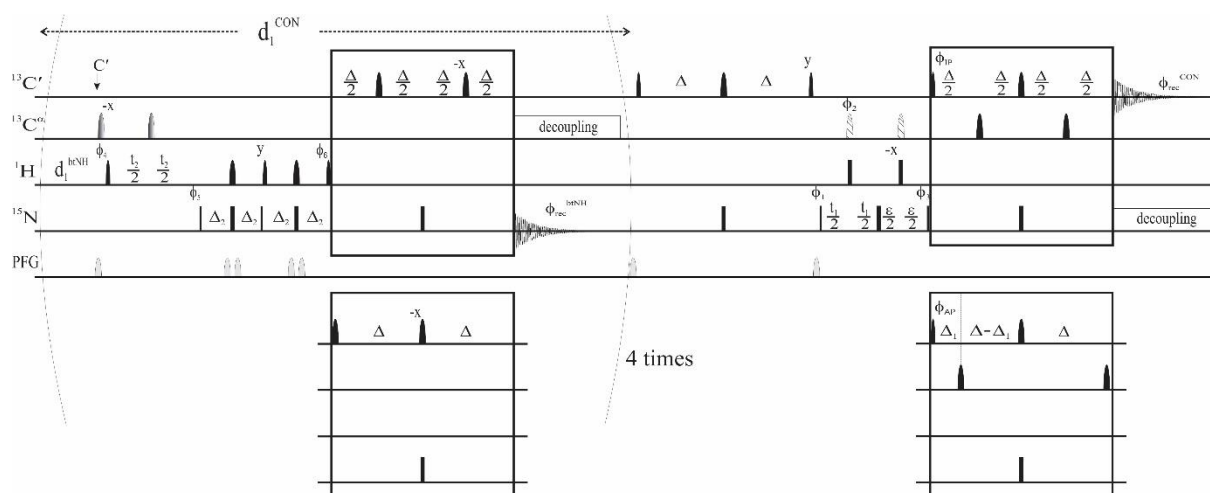
Figure S2. CON//btNH pulse sequence scheme.

Narrow and wide black rectangles represent $\pi/2$ and π non-selective pulses; narrow and wide black symbols represent $\pi/2$ and π band-selective pulses (Q5 and Q3 shapes of 300 and 231 μs respectively), hatched pulses are adiabatic inversion pulse on C' and C$^\alpha$ (smoothed Chirp 500 μs 80 kHz sweep and 20% smooting), the shaded shapes are Bip pulses.

The following phase cycling was employed for the CON//btNH: ϕ_1 = x, -x; ϕ_2 = 2(x), 2(-x); ϕ_3 = 4(x), 4(-x) and ϕ_{rec}^{CON}= x, -x, x, -x, -x, x, -x, x for the 2D-CON and ϕ_4 =y, -y, x, -x; ϕ_5 = y ϕ_6 = x and ϕ_{rec}^{btNH} = x, -x, -y, y for the 2D-btNH. Quadrature detection in the indirect dimension was achieved through the STATES-TPPI approach incrementing phase ϕ_1 for the CON and through the Echo/Antiecho approach by incrementing phase ϕ_4 and ϕ_5.

The length of the delays was: d_1= 2.35 s; Δ= 16.6 ms; Δ_1 = 4.5 ms; Δ_2= 2.7 ms; ε= $t_1(0)$ + duration of adiabatic pulse (500 μs).

Virtual decoupling of the C'-C$^\alpha$ coupling in the 2D-CON was achieved by acquiring for each increment both the IP and AP component of the signals. The same approach was also employed to achieve heteronuclear decoupling (C'-N) during acquisition of the ^{15}N FIDs in order to preserve the C' polarization for the subsequent CON experiment. Band selective C$^\alpha$ decoupling was achieved using a Q3 pulse in a P5M4 supercycle.
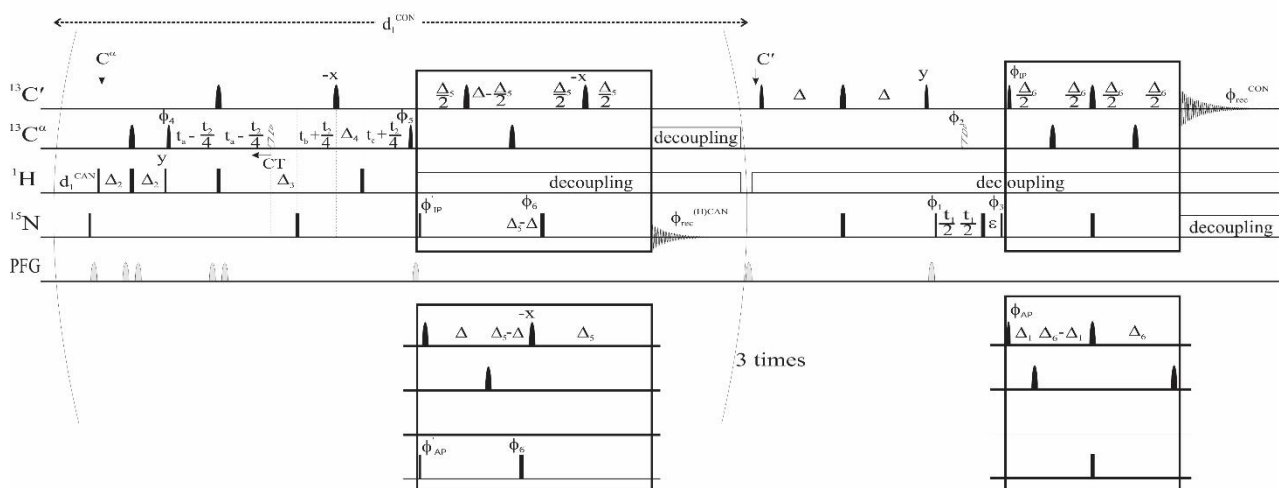
Figure S3. CON//(H)CAN pulse sequence scheme.

Narrow and wide black rectangles represent $\pi/2$ and π non-selective pulses; narrow and wide black symbols represent $\pi/2$ and π band-selective pulses (Q5 and Q3 shapes of 300 and 231 μs respectively), hatched pulses are adiabatic inversion pulse on C' and C^α (smoothed Chirp 500 μs 80 kHz sweep and 20% smooting).

The following phase cycling was employed for the CON//(H)CAN: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x)$, $4(-x)$; $\phi^{IP} = x$; $\phi^{AP} = -y$ and $\phi_{rec}^{CON} = x, -x, x, -x, -x, x, -x, x$ for the 2D-CON and $\phi_4 = x, -x$, $\phi_5 = 4(x), 4(-x)$ $\phi_6 = 8(x)\ 8(-x)$; $\phi'^{IP} = 2(x), 2(-x)$; $\phi'^{AP} = 2(y), 2(-y)$ and $\phi_{rec}^{(H)CAN} = x, -x, -x, x, -x, x, x, -x$ for the 2D-(H)CAN. Quadrature detection in the indirect dimension was achieved through STATES-TPPI approach incrementing phase ϕ_1 (CON) and ϕ_4 ((H)CAN).

The length of the delays was: $d_1= 2.5$ s; $\Delta = 12.4$ ms; $\Delta_1 = 4.5$ ms; $\Delta_2 = 1.7$ ms; $\Delta_3 = 1.9$ ms; $\Delta_4 = 1.4$ ms; $\Delta_5 = 4.2$ ms; $\Delta_6 = 16.6$ ms. t_a, t_b and t_c were used to achieve the constant time mode for the $^{13}C^\alpha$ indirect dimension.

Virtual decoupling of the C'-C^α coupling in the CON was achieved by acquiring for each increment both the IP and AP component of the signals. The same approach was also employed to achieve heteronuclear decoupling (C'-N) during acquisition of the ^{15}N FIDs in order to preserve the C' polarization for the subsequent CON experiment. Band selective C^α decoupling was achieved using a Q3 pulse in a P5M4 supercycle.

4

Pulse sequence CON//HN

```
;ut_con_fhsqchn
;avance-version

;Dataset 1 (F1)
;fhsqcf3gpph
;avance-version (12/01/11)
;2D H-1/X correlation via double inept transfer
;
;       F1(HN) -> F3(N,t1) -> F1(HN,t2)
;
;phase sensitive
;with decoupling during acquisition
;
;(S. Mori, C. Abeygunawardana, M. O'Neil-Johnson & P.C.M. van Zijl,
;   J. Magn. Reson. B 108, 94-98 (1995) )
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


;Dataset 2 (F2)
;CON
;2D sequence with
;   13C detected correlation for triple resonance using
;       inept transfer steps
;
;       F1(C=O) -> F3(N,t1) -> F1(C=O,t2)
;
;on/off resonance 13C pulses using shaped pulses
;phase sensitive (t1)
;using IPAP scheme for virtual decoupling
;(use parameterset C_CON_IASQ)
;
;W. Bermel, I. Bertini,L. Duma, I.C. Felli, L. Emsley,  R. Pierattelli,
;   P.R. Vasos, Angew. Chem. Int. Ed. 44, 3089-3092 (2005)
;(L. Duma, S. Hediger, A. Lesage & L. Emsley,
;   J. Magn. Reson. 164, 187-195 (2003) )
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


prosol relations=<triple>


#include <Avance.incl>
#include <Delay.incl>
#include <Grad.incl>


"p2=p1*2"
"p22=p21*2"
"d11=30m"
"d12=20u"
"d26=1s/(cnst4*4)"

"d22=4.5m"
"d23=16.6m"

"in31=inf1/2"
"in32=in2/2"
```

```
"d31=3u"
"d32=3u"


"DELTA=d19-p22/2"
"DELTA1=d26-p16-d16-p27*3-d19*5-p1*2/PI"
"DELTA2=d26-p16-d16-p27*2-p0-d19*5-de-8u"
"DELTA3=d31+larger(p2,p8)/2"
"DELTA4=p21*2/PI"
"DELTA5=d32*2+p8"
"DELTA6=d23/2-p14/2"
"DELTA7=d23-d22-p14"

"TAU=d26-p16-d16-4u"


"l0=1"

"l3=td1/2"


"spoff2=0"
"spoff3=0"
"spoff5=bf2*((cnst22-cnst21)/1000000)"
"spoff8=0"
"spoff13=bf2*((cnst26-cnst21)/1000000)"


1 4u ze1
  4u ze2

2 d11 do:f1 do:f3
  3m
3 12m
4 6m
5 d1

; fhsqc

  20u fq=cnst47(bf ppm):f3
  d12 pl1:f1 pl3:f3
  50u UNBLKGRAD

  (p1 ph11)
  4u
  p16:gp1
  d16
  TAU pl3:f3
  (center (p2 ph11) (p22 ph16):f3 )
  4u
  TAU
  p16:gp1
  d16
  (p1 ph12)

  4u
  p16:gp2
  d16
  (p8:sp13 ph17):f2
  4u

  (p21 ph13):f3
  DELTA3
  (p22 ph13):f3
  DELTA4
  d31

  (center (p2 ph15) (p8:sp13 ph11):f2 )
```

6

```
   d31
   DELTA4
   (p22 ph14):f3
   DELTA3
   (p21 ph14):f3

   4u
   p16:gp2
   d16

   (p1 ph17)
   p16:gp3
   d16 pl18:f1
   DELTA1
   p27*0.231 ph18
   d19*2
   p27*0.692 ph18
   d19*2
   p27*1.462 ph18
   DELTA
   (p22 ph11):f3
   DELTA
   p27*1.462 ph19
   d19*2
   p27*0.692 ph19
   d19*2
   p0*0.231 ph19
   4u
   p16:gp3
   d16
   4u BLKGRAD
   DELTA2 pl16:f3

   goscnp1 ph31 cpd3:f3

; con

   4u do:f3
   50u UNBLKGRAD

   p16:gp5
   d16

   20u pl19:f1 pl3:f3
   20u fq=cnst21(bf ppm):f2
   20u fq=cnst57(bf ppm):f3
   d12 cpds1:f1

   (p13:sp2 ph1):f2
   d23
   (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
   d23
   (p13:sp8 ph2):f2

   p16:gp4
   d16

   (p21 ph3):f3
   d32
   (p8:sp13 ph5):f2
   d32
   (p22 ph1):f3
   DELTA5
   (p21 ph4):f3

   if "l0 %2 == 1"
      {
      (p13:sp2 ph1):f2
      DELTA6
      (p14:sp5 ph1):f2
```

7

```
    DELTA6
    (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
    DELTA6
    (p14:sp5 ph1):f2
    DELTA6 pl16:f3
    }
  else
    {
    (p13:sp2 ph6):f2
    d22
    (p14:sp5 ph1):f2
    DELTA7
    (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
    DELTA6
    DELTA6 pl16:f3
    (p14:sp5 ph1):f2
    }

  4u BLKGRAD
  go2=2 ph30 cpd3:f3

  d11 do:f1 do:f3 wr2 #1 if2 #1 zd2

;exp_f2 ipap
  3m iu0
  lo to 3 times 2

  3m wr1 #0 if1 #0 zd1

;exp_f1 phase
  3m ip13
  3m ip16

;exp_f2 phase
  3m ip3
  lo to 4 times 2

;exp_f1 delay
  3m id31

;exp_f2 delay
  3m id32
  lo to 5 times l3

exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 0 0 2 2 2 2
ph5=0 0 2 2
ph6=3

ph11=0
ph12=1
ph13=0 2
ph14=0 0 0 0 2 2 2 2
ph15=0 0 2 2
ph16=0
ph17=2
ph18=1
ph19=3

ph30=0 2 0 2 2 0 2 0
ph31=0 2 0 2 2 0 2 0

;pl1 : f1 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl16: f3 channel - power level for CPD/BB decoupling
```

```
;pl18: f1 channel - power level for 3-9-19-pulse (watergate)
;pl19: f1 channel - power level for CPD/BB decoupling
;sp2 : f1 channel - shaped pulse  90 degree  (on resonance)
;sp3 : f1 channel - shaped pulse 180 degree  (on resonance)
;sp5 : f1 channel - shaped pulse 180 degree  (Ca off resonance)
;sp8 : f1 channel - shaped pulse  90 degree  (on resonance)
;                   for time reversed pulse
;sp13: f1 channel - shaped pulse 180 degree  (adiabatic)
;p0 : f1 channel -  90 degree pulse at pl18 for fine adjustment
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p8 : f1 channel - 180 degree shaped pulse for inversion (adiabatic)
;p13: f1 channel -  90 degree shaped pulse
;p14: f1 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse                          [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p27: f1 channel -  90 degree pulse at pl18
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                                [30 msec]
;d12: delay for power switching                         [20 usec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;d22: 1/(4J(COCa))                                      [4.5 msec]
;d23: 1/(4J(NCO))                                       [16.6 msec]
;d26: 1/(4J(NH))
;d31: incremented delay (2D, exp 1)                     [3 usec]
;d32: incremented delay (2D, exp 2)                     [3 usec]
;cnst4: = J(NH)
;cnst21: CO chemical shift (offset, in ppm)
;cnst22: Calpha chemical shift (offset, in ppm)
;cnst26: Call chemical shift (offset, in ppm)           [101 ppm]
;cnst47: N(H) chemical shift (offset, in ppm)
;cnst57: N chemical shift (offset, in ppm)
;o1p: CO chemical shift (cnst21)
;l0: flag to switch between inphase and antiphase
;l3 : loop for 2D experiment = td1/2
;inf1: 1/SW(N) = 2 * DW(N)
;in2 : 1/SW(N) = 2 * DW(N) (inf1 of exp 2, N all)
;in31: 1/(2 * SW(N)) =  DW(N)
;nd31: 2
;in32: 1/(2 * SW(N)) =  DW(N)
;nd32: 2
;ns: 8 * n
;ds: >= 32
;td1: number of experiments in F1
;       (number of experiments after IPAP processing)
;FnMODE: States-TPPI (or TPPI) in F1
;cpds1: decoupling according to sequence defined by cpdprg1
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd1: f1 channel - 90 degree pulse for decoupling sequence
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence

;for z-only gradients:
;gpz1: 50%
;gpz2: 80%
;gpz3: 31%
;gpz4: 19%
;gpz5: 60%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100

;use AU-program splitcomb [ipap 2] to process data

;$Id: $
```

Pulse sequence CON//btNH

```
;ut_con_n_btnh
;avance-version

;Dataset 1 (F1)
;c_con_iasq
;avance-version
;CON
;2D sequence with
;   13C detected correlation for triple resonance using
;       inept transfer steps
;
;       F1(C=O) -> F3(N,t1) -> F1(C=O,t2)
;
;on/off resonance 13C pulses using shaped pulses
;phase sensitive (t1)
;using IPAP scheme for virtual decoupling
;(use parameterset C_CON_IASQ)
;
;W. Bermel, I. Bertini,L. Duma, I.C. Felli, L. Emsley,  R. Pierattelli,
;   P.R. Vasos, Angew. Chem. Int. Ed. 44, 3089-3092 (2005)
;(L. Duma, S. Hediger, A. Lesage & L. Emsley,
;   J. Magn. Reson. 164, 187-195 (2003) )
;
;
;Dataset 2 (F2)
;dummy dataset
;
;
;Dataset 3 (F3)
;nb_hntrosy_f2ig
;avance-version (17/02/10)
;2D 15N shift correlation
;
;       F2(HN,t1) -> F3(N,t2)
;
;with refocussing of chemical shifts
;phase sensitive
;using semi-constant time in t1
;(use parameterset )
;
;(R. Schnieders, C. Richter, S. Warhaut, V. de Jesus, S. Keyhani,
;   E. Duchardt-Ferner, H. Keller, J. Wöhnert, L.T. Kuhn, A.L. Breeze,
;   W. Bermel, H. Schwalbe & B. Fürtig, ;   J. Biomol. NMR 69, 31–44 (2017) )
;(K. Takeuchi, H. Arthanari, I. Shimada & G. Wagner, J. Biomol. NMR 63, 1–9 (2015) )
;(K. Takeuchi, H. Arthanari, M. Imai & G. Wagner, J. Biomol. NMR 64, 143–151 (2016) )
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


prosol relations=<triple_c>


#include <Avance.incl>
#include <Delay.incl>
#include <Grad.incl>


"p22=p21*2"
"d11=30m"
"d12=20u"
"d26=1s/(cnst4*4)"

"d22=4.5m"
```

```
"d23=16.6m"


"p42=(bwfac42/(cnst55*cnst52*bf2))*1000000"
"spw42=plw2/((p42*90.0)/(p3*totrot42))*((p42*90.0)/(p3*totrot42))*(integfac42*integ-
fac42)"
"spoal42=0.5"
"spoff42=bf2*(cnst54/1000000)-o2"

"p43=(bwfac43/(cnst55*cnst53*bf2))*1000000"
"spw43=plw2/((p43*90.0)/(p3*totrot43))*((p43*90.0)/(p3*totrot43))*(integfac43*integ-
fac43)"
"spw44=plw2/((p43*90.0)/(p3*totrot44))*((p43*90.0)/(p3*totrot44))*(integfac44*integ-
fac44)"
"spoal43=0"
"spoal44=1"
"spoff43=bf2*(cnst54/1000000)-o2"
"spoff44=bf2*(cnst54/1000000)-o2"


"in31=inf1/2"
"in32=in2/2"

"d31=3u"
"d32=in32/2-p39/2"


"d60=d1-(d61+d59)*4-d58"


"DELTA=d31"
"DELTA1=d23/2-p12/2"
"DELTA2=d23-d22-p12"
"DELTA11=d26-larger(p22,p42)/2-p16-d16"
"DELTA12=d26-larger(p22,p42)/2-p16-d16-p21*2/PI"
"DELTA13=d26-larger(p22,p42)/2-p16-d16-de"
"DELTA14=d23/2-p12/2"


"l0=0"
"l3=l22/2"
"l10=1"
"l13=td1/(8*l22)"
"l30=0"


"spoff4=bf1*((cnst26-cnst22)/1000000)"
"spoff13=bf1*((cnst26-cnst21)/1000000)"
"spoff23=0"
"spoff24=0"
"spoff25=0"
"spoff26=bf1*((cnst22-cnst21)/1000000)"
"spoff27=bf1*((cnst21-cnst22)/1000000)"


1 4u ze1
  4u ze3
2 d11 do:f3
  3m
3 3m
4 3m
5 6m
6 9m
7 12.02m
8 3m

  d60
  50u UNBLKGRAD

goscnp2 ph31
```

```
   p16:gp2
   d16

;n_trhn

   d12 pl19:f1 pl3:f3
   20u fq=cnst22(bf ppm):f1

   "cnst31=(l30%2)*90"

   3m ip18+cnst31

11 d61

   (p39:sp4 ph17):f1

   if "l0 %2 == 1"
      {
      (p43:sp44 ph13+ph18):f2
      }
   else
      {
      (p43:sp44 ph14+ph18):f2
      }
   d32
   (p39:sp4 ph1+ph18):f1
   d32

   (p21 ph15+ph18):f3
   DELTA11
   p16:gp3
   d16
   4u
   (center (p42:sp42 ph1+ph18):f2 (p22 ph1+ph18):f3 )
   4u
   p16:gp3
   d16
   DELTA11
   (p21 ph1+ph18):f3

   (p43:sp43 ph2+ph18):f2
   DELTA12
   p16:gp4
   d16
   4u
   (center (p42:sp42 ph1+ph18):f2 (p22 ph1+ph18):f3 )
   4u
   p16:gp4
   d16
   DELTA13
   4u BLKGRAD
   (p43:sp44 ph16+ph18):f2

   if "l30 %2 == 0"
      {
      DELTA14
      (p12:sp27 ph1):f1
      DELTA14
      (p22 ph1):f3
      DELTA14
      (p12:sp27 ph17):f1
      DELTA14 pl19:f1
      }
   else
      {
      (p12:sp27 ph1):f1
      DELTA14
      DELTA14
      (p22 ph1):f3
```

```
      (p12:sp27 ph17):f1
      DELTA14
      DELTA14 pl19:f1
      }

  goscnp3 ph30 cpd1:f1

  4u do:f1
  lo to 11 times 4

  50u UNBLKGRAD

  p16:gp5
  d16

;con

  20u pl3:f3
  20u fq=cnst21(bf ppm):f1

  (p11:sp23 ph1)
  d23
  (center (p12:sp24 ph1) (p22 ph1):f3 )
  d23
  (p11:sp25 ph2)

  p16:gp1
  d16

  (p21 ph3):f3
  d31
  (center (p8:sp13 ph5):f1 (p44:sp3 ph1):f2 )
  d31
  (p22 ph1):f3
  DELTA
  (center (p8:sp13 ph1):f1 (p44:sp3 ph7):f2 )
  DELTA
  (p21 ph4):f3

  if "l10 %2 == 1"
      {
      (p11:sp23 ph1)
      DELTA1
      (p12:sp26 ph1)
      DELTA1
      (center (p12:sp24 ph1) (p22 ph1):f3 )
      DELTA1
      (p12:sp26 ph1)
      DELTA1 pl16:f3
      }
  else
      {
      (p11:sp23 ph6)
      d22
      (p12:sp26 ph1)
      DELTA2
      (center (p12:sp24 ph1) (p22 ph1):f3 )
      DELTA1
      DELTA1 pl16:f3
      (p12:sp26 ph1)
      }

  4u BLKGRAD

  go1=2 ph31 cpd3:f3

  d11 do:f3 wr1 #0 if1 #0 zd1
  3m wr2 #1 if2 #1 zd2

;exp_f1 ipap
```

```
  3m iu10
  lo to 3 times 2

;exp_f1 phase
  3m ip3
  lo to 4 times 2

;exp_f1 delay
  3m id31
  lo to 5 times l13

  3m wr3 #2 if3 #2 zd3

  3m iu30
  lo to 6 times 2

;exp_f2 phase
  3m ip15*2
  3m ip16*2
  3m iu0
  lo to 7 times 2

;exp_f2 delay
  3m id32
  3m
  20u
  3m
  3m
  lo to 8 times l3


exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 0 0 2 2 2 2
ph5=0 0 2 2
ph6=3
ph7=2
ph13=1 3 0 2
ph14=3 1 0 2
ph15=1
ph16=0
ph17=2
ph18=0
ph30=0 2 3 1
ph31=0 2 0 2 2 0 2 0


;pl1 : f1 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl16: f3 channel - power level for CPD/BB decoupling
;pl19: f1 channel - power level for CPD/BB decoupling
;sp3 : f2 channel - shaped pulse 180 degree (Bip720,50,20.1)
;sp4 : f1 channel - shaped pulse 180 degree (Bip720,100,10.1)
;sp13: f1 channel - shaped pulse 180 degree  (adiabatic)
;sp23: f1 channel - shaped pulse  90 degree  (on resonance)
;sp24: f1 channel - shaped pulse 180 degree  (on resonance)
;sp25: f1 channel - shaped pulse  90 degree  (on resonance)
;                     for time reversed pulse
;sp26: f1 channel - shaped pulse 180 degree  (Ca off resonance)
;sp27: f1 channel - shaped pulse 180 degree  (C=O off resonance)
;sp42: f2 channel - shaped pulse 180 degree (Reburp.1000)
;sp43: f2 channel - shaped pulse  90 degree (Eburp2.1000)
;sp44: f2 channel - shaped pulse  90 degree (Eburp2tr.1000)
;p8 : f1 channel - 180 degree shaped pulse for inversion (adiabatic)
;p11: f1 channel -  90 degree shaped pulse
;p12: f1 channel - 180 degree shaped pulse
```

```
;p16: homospoil/gradient pulse                          [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p39: f1 channel - 180 degree shaped pulse for refocussing
;                 Bip720,100,10.1 (160us at 600.13 MHz)
;p42: f2 channel - 180 degree shaped pulse for refocussing
;                 Reburp.1000               (1.4ms at 600.13 MHz)
;p43: f2 channel -  90 degree shaped pulse for excitation
;                 Eburp2.1000/Eburp2tr.1000 (1.7ms at 600.13 MHz)
;p44: f2 channel - 180 degree shaped pulse for refocussing
;                 Bip720,50,20.1          (200us at 600.13 MHz)
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                                 [30 msec]
;d12: delay for power switching                          [20 usec]
;d16: delay for homospoil/gradient recovery
;d22: 1/(4J(COCa))                                       [4.5 msec]
;d23: 1/(4J(NCO))                                        [16.6 msec]
;d26: 1/(4J)NH
;d31: incremented delay (2D, exp 1))                     [3 usec]
;d32: incremented delay (2D, exp 3)): in32/2-p39/2
;d58: = AQ of exp 2 (1H)
;d59: = AQ of exp 3 (15N)
;d60: relaxation delay as executed
;d61: relaxation delay of exp 3
;cnst4: = J(NH)
;cnst21: CO chemical shift (offset, in ppm)
;cnst22: Calpha chemical shift (offset, in ppm)
;cnst26: Call chemical shift (offset, in ppm)            [101 ppm]
;cnst43: compensation of chemical shift evolution during p43
;          Eburp2.1000: 0.69
;cnst52: scaling factor for p42 to compensate for transition region
;          Reburp.1000: 1.426
;cnst53: scaling factor for p43 to compensate for transition region
;          Eburp2.1000: 1.000
;cnst54: H(N) chemical shift (offset, in ppm)
;cnst55: H(N) bandwidth (in ppm)
;o1p: CO chemical shift (cnst21)
;l0 : flag to switch between odd and even increments (exp 3)
;l3 : loop for 2D exp 1 = l22/2
;l13: loop for 2D exp 3 = td1/(8*l22)
;l10: flag to switch between inphase and antiphase (exp 1)
;l22: number of experiments in F1 for exp 3 (after IPAP processing)
;l30: flag to switch between inphase and antiphase (exp 3)
;inf1: 1/SW(N) = 2 * DW(N)
;in2 : 1/SW(H) = 2 * DW(H) (inf1 of exp 3)
;in31: 1/(2 * SW(N)) =  DW(N)
;nd31: 2
;in32: 1/(2 * SW(H)) = DW(H)
;nd32: 2
;ns: 8 * n
;ds: >= 32
;td1: number of experiments in F1
;       (number of experiments after IPAP processing)
;FnMODE: States-TPPI (or TPPI) in F1
;cpd1: decoupling according to sequence defined by cpdprg1: p5m4sp180
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd1: f1 channel - 180 degree pulse for decoupling sequence
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence
;cpdprg1: p5m4sp180


;for z-only gradients:
;gpz1: 50%
;gpz2: 70%
;gpz3: 31%
;gpz4: 19%
;gpz5: 45%


;use gradient files:
;gpnam1: SMSQ10.100
```

```
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100


;use AU-program splitcomb [ipap 2] to process data

;Processing (exp 3)

;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1


;$Id: $
```

Pulse sequence CON//(H)CAN

```
;ut_con_n_(h)can
;avance-version

;Dataset 1 (F1)
;n_hcan.2
;avance-version
;2D sequence with
;   15N detected correlation for triple resonance using
;      inept transfer steps
;
;      F3(Ha) -> F2(Ca,t1) -> F1(N,t2)
;
;on/off resonance 13C pulses using shaped pulses
;phase sensitive (t1)
;using constant time in t1
;(use parameterset )
;
;M. Gal, K. A. Edmonds, A. G. Milbradt, K. Takeuchi & G. Wagner,
;   J Biomol NMR 51, 497-504 (2011)
;(K. Takeuchi, G. Heffron, Z.-Y. J. Sun, D. P. Frueh & G. Wagner,
;   J Biomol NMR 47, 271-282 (2010))
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


;Dataset 2 (F2)
;c_con_iasq
;avance-version
;CON
;2D sequence with
;   13C detected correlation for triple resonance using
;      inept transfer steps
;
;      F2(C=O) -> F1(N,t1) -> F2(C=O,t2)
;
;on/off resonance 13C pulses using shaped pulses
;phase sensitive (t1)
;using IPAP scheme for virtual decoupling
;(use parameterset )
;
;W. Bermel, I. Bertini,L. Duma, I.C. Felli, L. Emsley,  R. Pierattelli,
;   P.R. Vasos, Angew. Chem. Int. Ed. 44, 3089-3092 (2005)
;(L. Duma, S. Hediger, A. Lesage & L. Emsley,
;   J. Magn. Reson. 164, 187-195 (2003) )
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


prosol relations=<triple>


#include <Avance.incl>
#include <Delay.incl>
#include <Grad.incl>


"p2=p1*2"
"p22=p21*2"
"d3=1s/(cnst2*cnst12)"
```

17

```
"d4=1s/(cnst2*4)"
"d11=30m"
"d12=20u"

"d22=4.5m"
"d23=12.4m"
"d25=16.6m"
"d27=14.3m"


"d32=3u"
"d31=d27/2-d3-p22/2"
"d41=d27/2-larger(p14,p22)/2"
"d51=d23-d27/2-p14/2-p2/2"


"in31=inf1/4"
"in32=in2/2"

"in41=in31"
"in51=in31"

"td1=tdmax(td1,d41*2,in41)"


"d1=(d60+d58)*3 +d59+50m"


"DELTA1=d27-d23-p2/2"
"DELTA2=d3-p14/2-p22/2"
"DELTA3=d25/2"
"DELTA4=d23-d25/2-p14"
"DELTA5=d25-d23"
"DELTA6=d25/2-d12+p1*2/PI-de-4u"
"DELTA7=d32*2+p8"
"DELTA8=d23/2-p14/2"
"DELTA9=d23-d22-p14"


"spoff2=0"
"spoff3=0"
"spoff5=bf2*((cnst21-cnst22)/1000000)"
"spoff7=bf2*((cnst22-cnst21)/1000000)"
"spoff8=0"
"spoff13=bf2*((cnst26-cnst21)/1000000)"
"spoff23=bf2*((cnst23-cnst22)/1000000)"

"l0=0"
"l10=0"

"l3=td1/4"


"acqt0=0"
baseopt_echo


1 4u ze1
  4u ze2
  d1

2 d11 do:f1 do:f3
  3m
3 3m
4 9m
5 3m
6 9m
7 3m

;n_hcan
```

```
8 d11 do:f2 do:f3
  d60

  50u UNBLKGRAD
  4u pl1:f1 pl3:f3

  (p1 ph1):f1
  p16:gp1
  d16 fq=cnst22(bf ppm):f2

  (p21 ph1):f3
  d4
  (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
  d4
  (p21 ph2):f3

  (p13:sp2 ph3):f2
  d41
  (center (p14:sp5 ph1):f2 (p22 ph1):f3 )
  d41
  (p14:sp23 ph1):f2
  DELTA1
  (p2 ph1):f1
  d51
  (p14:sp5 ph10):f2
  DELTA2
  (p22 ph1):f3
  d31
  (p13:sp8 ph4):f2

  4u
  p16:gp2
  d16 pl16:f3
  20u cpd3:f3

  if "l0 %2 == 0"
     {
     (p1 ph5):f1
     DELTA3
     (p14:sp5 ph1):f2
     DELTA4
     (p14:sp3 ph1):f2
     DELTA5
     (p2 ph6):f1
     DELTA3
     (p14:sp5 ph10):f2
     }
  else
     {
     (p1 ph7):f1
     (p14:sp5 ph1):f2
     DELTA3
     DELTA4
     (p14:sp3 ph1):f2
     DELTA5
     (p2 ph6):f1
     (p14:sp5 ph10):f2
     DELTA3
     }

  DELTA6
  d12 pl28:f2
  4u BLKGRAD

  goscnp1 ph31 cpd2:f2
  lo to 8 times 3


;c_con
```

```
4u do:f2 do:f3
50m
50u UNBLKGRAD

p16:gp3
d16 fq=cnst21(bf ppm):f2

4u pl1:f1 pl16:f3
4u cpd3:f3

(p13:sp2 ph1):f2
d23
(center (p2 ph1):f1 (p14:sp3 ph1):f2 )
d23
(p13:sp8 ph2):f2

p16:gp4
d16

(p1 ph13):f1
d32
(p8:sp13 ph15):f2
d32
(p2 ph1):f1
DELTA7
(p1 ph14):f1

if "l10 %2 == 0"
    {
    (p13:sp2 ph1):f2
    DELTA8
    (p14:sp7 ph1):f2
    DELTA8
    (center (p2 ph1):f1 (p14:sp3 ph1):f2 )
    DELTA8
    (p14:sp7 ph1):f2
    DELTA8 pl19:f1
    }
  else
    {
    (p13:sp2 ph16):f2
    d22
    (p14:sp7 ph1):f2
    DELTA9
    (center (p2 ph1):f1 (p14:sp3 ph1):f2 )
    DELTA8
    DELTA8 pl19:f1
    (p14:sp7 ph1):f2
    }

4u BLKGRAD

go2=2 ph30 cpd1:f1

d11 do:f1 do:f3 wr2 #1 if2 #1 zd2

;exp_f2 ipap
3m iu10
lo to 3 times 2

;exp_f2 phase
3m dp13
lo to 4 times 2

3m wr1 #0 if1 #0 zd1

;exp_f1 ipap
3m iu0
```

```
;exp_f2 delay
  3m id32
  lo to 5 times 2

;exp_f1 phase
  3m ip3
  lo to 6 times 2

;exp_f1 delay
  3m id31
  3m dd41
  3m id51

  lo to 7 times l3

exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 0 0 2 2 2 2
ph5=0 0 2 2
ph6=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph7=1 1 3 3
ph10=2

ph13=0 2
ph14=0 0 0 0 2 2 2 2
ph15=0 0 2 2
ph16=3

ph29=0
ph30=0 2 0 2 2 0 2 0
ph31=0 2 2 0 2 0 0 2


;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl16: f3 channel - power level for CPD/BB decoupling
;pl19: f1 channel - power level for CPD/BB decoupling
;pl28: f2 channel - power level for CPD/BB decoupling
;sp2: f2 channel - shaped pulse  90 degree  (on resonance)
;sp3: f2 channel - shaped pulse 180 degree  (on resonance)
;sp5: f2 channel - shaped pulse 180 degree  (C=O off resonance)
;sp7: f2 channel - shaped pulse 180 degree  (Ca off resonance)
;sp8: f2 channel - shaped pulse  90 degree  (on resonance)
;                    for time reversed pulse
;sp13: f2 channel - shaped pulse 180 degree  (adiabatic)
;sp23: f2 channel - shaped pulse 180 degree  (Cali off resonance)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p8 : f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p13: f2 channel -  90 degree shaped pulse
;p14: f2 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse                    [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;d1 : relaxation delay as executed; 1-5 * T1
;d3 : 1/(n*J(CH)), n = cnst12
;d4 : 1/(4J(CH))
;d11: delay for disk I/O                          [30 msec]
;d12: delay for power switching                   [20 usec]
;d16: delay for homospoil/gradient recovery
;d22: 1/(4J(COCa))                                [4.5 msec]
;d23: 1/(4J(NCO))                                 [12.4 msec]
;d25: 1/(4J'(NCO))                                [16.6 msec]
;d27: 1/(2J(CaCb))                                [14.3 msec]
;d31 : incremented delay (2D, exp 1) = d27/2-d3-p22/2
```

```
;d32 : incremented delay (2D, exp 2)                    [3 usec]
;d41: decremented delay (2D, exp 1) = d27/2-larger(p14,p22)/2
;d51: incremented delay (2D) = d23-d27/2-p14/2-p2/2
;d58: = AQ of exp 1 (15N)
;d59: = AQ of exp 2 (13C)
;d60: relaxation delay for exp 1
;cnst2: = J(CH)                                         [145 Hz]
;cnst12: for multiplicity selection = 4 for all but Gly, 5 including Gly
;cnst21: CO chemical shift (offset, in ppm)
;cnst22: Calpha chemical shift (offset, in ppm)
;cnst23: Caliphatic chemical shift (offset, in ppm)
;cnst26: Call chemical shift (offset, in ppm)          [101 ppm]
;o1p: CO chemical shift (cnst21)
;l0 : flag to switch between inphase and antiphase (exp 1)
;l3 : loop for 2D experiment = td1/4
;l10: flag to switch between inphase and antiphase (exp 2)
;inf1: 1/SW(Ca) = 2 * DW(Ca)
;in2 : 1/SW(N) = 2 * DW(N) (inf1 of exp 2)
;in31: 1/(4 * SW(Ca)) = (1/2) DW(Ca)
;nd31: 4
;in32: 1/(2 * SW(N)) = DW(N)
;nd32: 2
;in42: = in32
;in52: = in32
;ns: 4 * n
;ds: >= 32
;td1: number of experiments in F1
;      (number of experiments after IPAP processing)
;FnMODE: States-TPPI (or TPPI) in F1
;cpd1: decoupling according to sequence defined by cpdprg1
;cpd2: decoupling according to sequence defined by cpdprg2: p5m4sp180
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd1: f1 channel - 90 degree pulse for decoupling sequence
;pcpd2: f2 channel - 180 degree pulse for decoupling sequence
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence

;for z-only gradients:
;gpz1: 50%
;gpz2: 40%
;gpz3: 60%
;gpz4: 30%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100

;use AU-program splitcomb [ipap 2] to process data

;$Id: $
```

Chapter 3. Structural and functional studies of IDPs by NMR

The number of structurally and functionally characterized IDPs is growing rapidly. Despite their abundance and their multiple biological activities, disordered proteins were ignored for a long time, as the structure-function paradigm dominated the scientific scene. It was only in the last two decades that disordered proteins started to be studied more systematically, when more than 15000 proteins in the Swiss Protein (SwissProt) database were predicted to contain intrinsically disordered regions (IDRs) of at least 40 consecutive amino acids.[92]

In the following paragraphs, different examples of application of NMR for IDPs investigation are reported. The large number of information provided by NMR spectroscopy allowed us to study the interaction between a disordered protein and a potential drug (Chapter 3.1), to characterize a large viral IDP (Chapter 3.2), as well as the interaction between two IDPs (Chapter 3.4). NMR spectroscopy can have a role also in the characterization of the processes leading to liquid-liquid phase transition, that only recently have started to be investigated (Chapter 3.3).

3.1 IDPs and small molecules

Due to its relevant biological importance, α-synuclein has been widely studied. This typical intrinsically disordered protein of 140 amino acids is often used as a standard for the development of NMR experiments tailored for the study of IDPs thanks to the efficient expression and purification protocol, the high solubility and moderate size and the acceptable sample stability (Chapter 2).

Despite the high level of disorder, α-synuclein can adopt a number of different conformational states depending on conditions and cofactors.[93] These include the helical membrane-bound form, a partially-folded state that is a key intermediate in aggregation and fibrillation, various oligomeric species, and fibrillary and amorphous aggregates. The widespread aggregation of α-synuclein to form proteinaceus inclusions, called Lewy bodies together with the progressive dopaminergic neuronal cell loss in *the substantia nigra pars compacta* are the neuropathological hallmarks of Parkinson's disease (PD), one of the most renown disease related to IDPs.[94,95]

The implication of α-synuclein in PD pathogenesis stimulated research for its *druggability*, with extensive screenings for small molecules capable to alter its aggregation propensity. Dopamine and related catecholamine were among the first molecules found to prevent α-synuclein fibril formation.[96]

NMR spectroscopy has a wide range of applications in the pharmaceutical industry, most of which can be broadly categorized as related to drug design.[97–99] As already described in the introduction, NMR allows the study of biomolecules at atomic resolution at physiological condition and it is the election technique for IDPs.

Thanks to the available characterization of α-synuclein it is possible to exploit NMR spectroscopy to investigate the mechanism of action at the molecular level of potential drugs such as NDGA. Exploiting very simple 2D experiments, we were able to follow the aggregation process directly in the NMR tube by monitoring the spectra over time. Selective chemical shifts perturbation helped identifying potential binding site, while variation in intensity of the signals suggested variation of flexibility of parts of the protein, providing information on the structural conformation. These results were part of a large international effort to establish the potential of NDGA as a drug as reported in the following article.

3.1.1 Article: Cyclized NDGA modifies dynamic α-synuclein monomers preventing aggregation and toxicity

SCIENTIFIC REPORTS

OPEN

Cyclized NDGA modifies dynamic α-synuclein monomers preventing aggregation and toxicity

Malcolm J. Daniels[1], J. Brucker Nourse Jr.[2], Hanna Kim[2], Valerio Sainati[3], Marco Schiavina[3], Maria Grazia Murrali[3], Buyan Pan[4], John J. Ferrie[4], Conor M. Haney[4], Rani Moons[5], Neal S. Gould[6], Antonino Natalello [7], Rita Grandori[7], Frank Sobott[8,9,10], E. James Petersson[4], Elizabeth Rhoades[4], Roberta Pierattelli [3], Isabella Felli [3], Vladimir N. Uversky [11,12], Kim A. Caldwell[2], Guy A. Caldwell[2], Edward S. Krol[13] & Harry Ischiropoulos[1,6,14]

Growing evidence implicates α-synuclein aggregation as a key driver of neurodegeneration in Parkinson's disease (PD) and other neurodegenerative disorders. Herein, the molecular and structural mechanisms of inhibiting α-synuclein aggregation by novel analogs of nordihydroguaiaretic acid (NDGA), a phenolic dibenzenediol lignan, were explored using an array of biochemical and biophysical methodologies. NDGA analogs induced modest, progressive compaction of monomeric α-synuclein, preventing aggregation into amyloid-like fibrils. This conformational remodeling preserved the dynamic adoption of α-helical conformations, which are essential for physiological membrane interactions. Oxidation-dependent NDGA cyclization was required for the interaction with monomeric α-synuclein. NDGA analog-pretreated α-synuclein did not aggregate even without NDGA-analogs in the aggregation mixture. Strikingly, NDGA-pretreated α-synuclein suppressed aggregation of naïve untreated aggregation-competent monomeric α-synuclein. Further, cyclized NDGA reduced α-synuclein-driven neurodegeneration in *Caenorhabditis elegans*. The cyclized NDGA analogs may serve as a platform for the development of small molecules that stabilize aggregation-resistant α-synuclein monomers without interfering with functional conformations yielding potential therapies for PD and related disorders.

Parkinson's disease (PD) is an age-related neurodegenerative disorder characterized by a progressive motor phenotype including tremors, rigidity, and bradykinesia. These symptoms are driven primarily by loss of dopamine-producing neurons in the *substantia nigra pars compacta*. Lewy bodies, intracellular proteinaceous inclusions, are the histopathological hallmark of PD. Immunohistochemical analysis of Lewy bodies revealed aggregated forms of α-synuclein, a 140 amino acid protein, as a major component[1,2]. Mutations, duplications, and triplications of the gene encoding α-synuclein cause dominantly-inherited familial forms of PD[3–6]. Furthermore, aggregation of wild type α-synuclein is involved in the pathogenesis of a diverse group of neurodegenerative diseases[1,7–14].

[1]Pharmacology Graduate Group, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA. [2]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35487, USA. [3]CERM and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Florence, 50019, Italy. [4]Department of Chemistry, University of Pennsylvania, Philadelphia, PA, 19104, USA. [5]Department of Chemistry, University of Antwerp, Antwerp, Belgium. [6]Department of Pediatrics, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA, 19104, USA. [7]Department of Biotechnology and Biosciences, University of Milan-Bicocca, Milan, Italy. [8]Biomolecular & Analytical Mass Spectrometry, Antwerp University, Antwerp, Belgium. [9]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom. [10]School of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom. [11]Department of Molecular Medicine and Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA. [12]Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, 142292, Russian Federation. [13]College of Pharmacy & Nutrition, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [14]Children's Hospital of Philadelphia Research Institute and Systems Pharmacology and Translational Therapeutics, the Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA. Correspondence and requests for materials should be addressed to H.I. (email: ischirop@pennmedicine.upenn.edu)

The pathological and genetic evidence implicating α-synuclein in neurodegeneration has sparked numerous studies of the normal function of α-synuclein and its role in disease pathogenesis. α-Synuclein has been implicated in synaptic function[15]. It preferentially binds lipid membranes with high curvature[16] and may modulate neurotransmitter release and synaptic function by affecting vesicular dynamics[17-21].

Repeated studies across various animal models have shown that expression of mutant α-synuclein with altered aggregation kinetics causes neurodegeneration[22,23]. Likewise, induction of α-synuclein aggregation in wildtype animals by seeding with α-synuclein aggregates—isolated from PD Lewy bodies[24] or generated *in vitro*[25-28]—induces progressive neurodegeneration. Indeed, this propagation of α-synuclein aggregation is not only observed in animal models, but in dopaminergic tissue grafted into PD patients[29]. α-Synuclein aggregation also occurs in animal models of dopamine neuron degeneration induced by oxidative chemical insult[30,31].

The implication of α-synuclein in PD pathogenesis stimulated several screens for small molecules that alter its aggregation. Dopamine and related catecholamines were among the first molecules found to prevent α-synuclein fibril formation[32]. Oxidation of vicinal hydroxyls in dopamine, and related phenols, induces formation of soluble α-synuclein oligomers that do not incorporate into fibrils[33-35]. Recently, dopamine-induced α-synuclein oligomers were shown to contribute to neurodegeneration in mice[36], but the precise mechanism of oligomer toxicity remains a subject of debate[37]. Dopamine oxidation may also contribute to neurodegeneration through modification of glucocerebrosidase and consequential lysosomal dysregulation[38].

Ongoing research has identified many other inhibitors of α-synuclein aggregation. Many of these inhibitors stabilize multimers or oligomers through direct interaction with α-synuclein (e.g. various phenols, catechols, and flavonoids[35,39-45], Anle138b[46,47], rifampicin[48]). A well-studied member of this groups is epigallocatechin gallate (EGCG), a polyphenol sharing the vicinal hydroxyls implicated in dopamine's interaction with α-synuclein. Despite their chemical similarities, dopamine and EGCG have divergent effects on the structure of α-synuclein[49]. In fact, while dopamine may induce toxic oligomers, adding EGCG during cell-free α-synuclein aggregation produces species less toxic to cells[39]. Other aggregation inhibitors alter α-synuclein-protein or -lipid interactions to prevent aggregation (e.g. NPT100-18A[50], squalamine[51], PcTS[52,53], Hsp70[54]). Recent studies have examined small molecules that directly stabilize α-synuclein monomers (e.g. BIOD303[55], nortriptyline[56], CLR01[57,58]). However, it remains unknown whether these small molecules perturb α-synuclein's lipid interactions, which are directly implicated in its role in neurotransmitter release[59-61].

In this study we employed nordihydroguaiaretic acid (NDGA), a phenolic dibenzenediol lignan that inhibits α-synuclein aggregation, to examine the mechanisms of α-synuclein aggregation inhibitors and evaluate its potential utility as a chemical platform for the development of novel aggregation inhibitors[40,62]. Recent work has exhaustively characterized the products and kinetics of NDGA oxidation as well as several novel analogs NDGA and several novel analogs[63]. We provide evidence that cyclized NDGA, formed during oxidation, interacts with α-synuclein to produce modified monomers. We demonstrate that these aggregation-resistant monomers retain their capacity to interact with phospholipid membranes and that they inhibit aggregation of untreated α-synuclein. Further, we demonstrate for the first time that cyclized NDGA reduces α-synuclein-driven neurodegeneration in a relevant animal model. Generating α-synuclein species that retain native structural dynamics while exerting a dominant-negative effect on α-synuclein aggregation and neurotoxicity represents a new paradigm for intervening in PD and related disorders.

Results

NDGA inhibits both primary and secondary nucleation-mediated α-synuclein aggregation.
Two established assays, measurement of insoluble protein and Thioflavin-T binding, quantified the effect of NDGA on the aggregation of monomeric α-synuclein. NDGA caused concentration-dependent inhibition of insoluble α-synuclein accumulation and amyloid-like aggregate formation (Figs 1A,B and S1). Maximal inhibition was achieved at stoichiometric, equimolar concentrations of NDGA to monomeric α-synuclein. To confirm the Thioflavin-T findings, the secondary structure of aggregates was quantified by circular dichroism (CD). NDGA caused a dose-dependent decrease in the β-sheet content of the α-synuclein aggregates, as indicated by molar ellipticity decreases at 200 nm and increases at 220 nm (Fig. 1C). Finally, transmission electron microscopy was used to image the α-synuclein aggregates produced in the presence or absence of NDGA. Typical α-synuclein fibrils were observed in the absence of NDGA, but not after aggregation with equimolar NDGA (Fig. 1D). These data demonstrate that stoichiometric concentrations of NDGA inhibit the formation of amyloid-like α-synuclein fibrils.

The preceding experiments examined the effect of NDGA on a typical α-synuclein aggregation assay, in which primary nucleation is the rate-limiting step for the formation of fibrils. Aggregation of α-synuclein can also proceed by a secondary nucleation surface-catalyzed formation of fibrils. To test the effect of NDGA on secondary nucleation-dependent aggregation, α-synuclein fibrils (representing 5% of total α-synuclein) were added with monomeric α-synuclein. Equimolar NDGA (molecule to monomeric α-synuclein) inhibited the formation of insoluble α-synuclein aggregates in the presence of fibril seeds (Fig. 1E). Together, these findings demonstrate that stoichiometric NDGA prevents both primary and secondary nucleation dependent aggregation of α-synuclein.

Interaction with α-synuclein requires oxidation and cyclization of NDGA.
We used N-acetylcysteine (NAC), an electron donor, to inhibit NDGA oxidation in the presence of α-synuclein. NAC was included in a typical α-synuclein aggregation mixture at 20:1 molar excess to NDGA. After aggregation for 3 days, levels of α-synuclein aggregation were measured by the established solubility assay. mNDGA, a NDGA analog incapable of oxidation and cyclization was included as a negative control (Supplementary Fig. S2). The addition of NAC allowed α-synuclein to aggregate in the presence of NDGA, suggesting that oxidation is necessary for NDGA to inhibit aggregation (Fig. 2A,B). The inclusion of NAC also reduced levels of quinone-containing
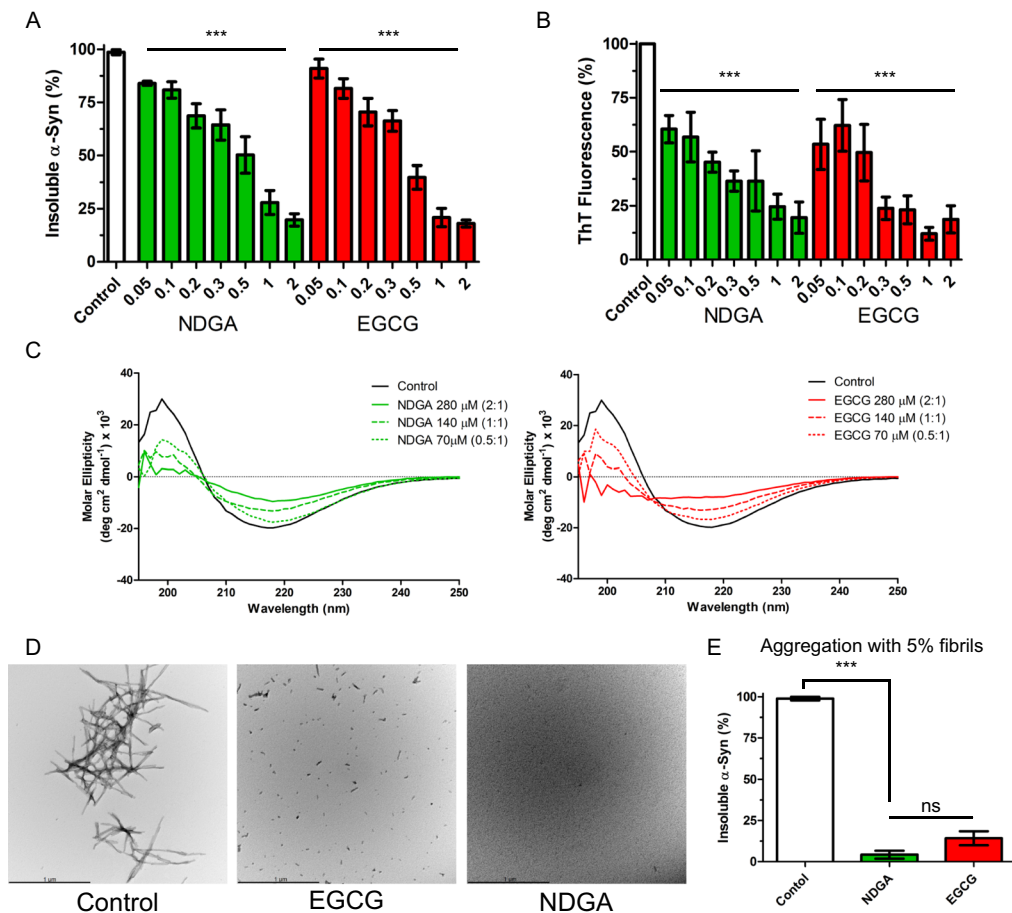
Figure 1. NDGA inhibits recombinant human α-synuclein aggregation. (**A**) Insoluble α-synuclein present after 7 days aggregation was reduced by NDGA and EGCG in a dose-dependent fashion as compared to solvent control. Recombinant human wildtype α-synuclein (138 μM) was aggregated for 7 days in the presence of EGCG or NDGA at the indicated molar ratios. After aggregation, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry (n = 3–5, ***p < 0.001). (**B**) Formation of amyloid α-synuclein fibrils quantified by Thioflavin-T was reduced by NDGA and EGCG in a dose-dependent fashion as compared to control after 7 days aggregation (n = 3–5, ***p < 0.001). (**C**) α-Synuclein beta-sheet secondary structure was reduced by EGCG and NDGA in a dose dependent fashion. Recombinant human wildtype α-synuclein (138 μM) was aggregated for 7 days in the presence of EGCG or NDGA at the indicated molar ratios. Secondary structure was quantified by circular dichroism. (**D**) Transmission electron microscopy images of α-synuclein aggregates after 3 days aggregation with small molecules at 1:1 molar ratio. (**E**) Insoluble α-synuclein present after 7 days aggregation in the presence of a 5% fibril seed was reduced by EGCG and NDGA. EGCG or NDGA were present at a 1:1 molar ratio. After aggregation, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry (n = 3, ***p < 0.001).

NDGA oxidation products detected by near-infrared fluorescence (nIRF), a technique developed to study the oxidation-dependent interaction between α-synuclein and dopamine[64] (Fig. 2C). The correlation of increased aggregation and decreased quinone nIRF suggests a relationship between NDGA oxidation and inhibition of α-synuclein aggregation. Separately, catalase, which rapidly converts H_2O_2 to water and hydrogen gas, was included in the aggregation assay to determine whether H_2O_2 that might be produced during NDGA oxidation mediated the effects on aggregation. Inclusion of catalase had no effect on α-synuclein aggregation in the presence of NDGA nor on quinone nIRF, indicating that the effect is not H_2O_2-dependent (Figs 2A and S3). These data indicate that inhibition of α-synuclein aggregation by NDGA and formation of quinone-modified α-synuclein is oxidation-dependent but not mediated by peroxide chemistry.

Previous work has identified the majority product of NDGA oxidation as a cyclized form[63]. During that study, novel NDGA analogs with defined oxidation pathways were generated. Here, these analogs were used to examine of the role of cyclization in NDGA's oxidation-dependent inhibition of α-synuclein aggregation. We employed NDGA and two such analogs: NDGA-1, which oxidizes with comparable kinetics but does not cyclize due to unilateral hydroxyl substitution, and NDGA-5, which cyclizes more readily than NDGA due to reduced steric
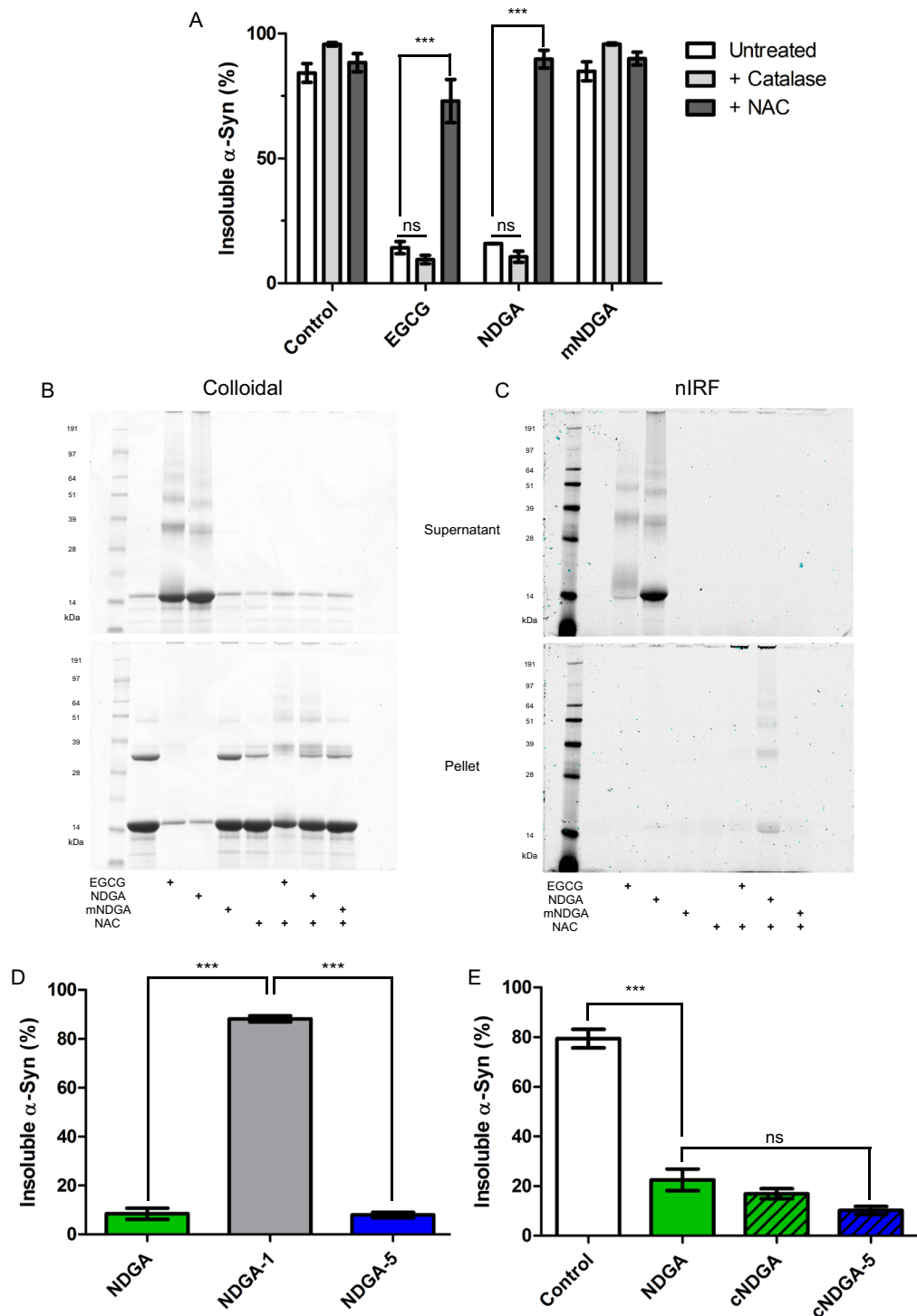
Figure 2. Interaction between NDGA and α-synuclein requires NDGA oxidation and cyclization. (**A**) NDGA treatment did not affect α-synuclein aggregation in the presence of N-acetylcysteine (NAC). α-Synuclein was aggregated for 3 days in the presence of 1:1 small molecules. After aggregation, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry. NAC was added at 20x molar excess to small molecule and catalase was added equal to 5% of protein, providing excess hydrogen peroxide decomposition capacity (n = 3–5, ***p < 0.001). (**B**) Colloidal staining of representative gels showed the formation of insoluble α-synuclein aggregates in the presence of NDGA and NAC. (**C**) Near-infrared fluorescent imaging of the same gels before colloidal staining showed a reduction of quinone-dependent fluorescence in the presence of NAC. (**D**) α-Synuclein did not aggregate in the presence of cyclizable analogs, NDGA and NDGA-5, but did with non-cyclizable NDGA-1 (n = 3, ***p < 0.001). (**E**) α-synuclein did not aggregate in the presence of cyclized cNDGA and cNDGA-5 (n = 3). α-Synuclein was aggregated for

3 days in the presence of 1:1 small molecules. After aggregation, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry.

hindrance in the crosslinking region (Supplementary Fig. S2). α-Synuclein was aggregated for 3 days in the presence of equimolar NDGA analogs, and aggregation was quantified by the solubility assay. Both cyclizing analogs, NDGA and NDGA-5, inhibited α-synuclein aggregation, while non-cyclizing NDGA-1 did not (Fig. 2D). The products of the above experiment where also analyzed by nIRF. NDGA and NDGA-5-treated α-synuclein exhibited quinone nIRF signals, while NDGA-1-treated α-synuclein did not (Supplementary Fig. S4). These results indicate that cyclized forms of NDGA, or the process of cyclization, are required for inhibition of α-synuclein aggregation and formation of quinone-modified α-synuclein. To address this uncertainty, α-synuclein was aggregated in the presence of cNDGA and cNDGA-5, the purified cyclized forms of NDGA and NDGA-5, respectively (Supplementary Fig. S2). Using the solubility assay, both cNDGA and cNDGA-5 inhibited α-synuclein aggregation as potently as NDGA (Figs 2E and S5). This indicates that NDGA's inhibition of α-synuclein aggregation is mediated by interaction of the protein with cyclized oxidation products.

NDGA induces modest, progressive α-synuclein compaction without preventing dynamic adoption of α-helical conformation.

During analysis of the interaction between NDGA and α-synuclein, equimolar incubation consistently produced nIRF-positive quinone-modified α-synuclein (Supplementary Fig. S6). Based on this observation, NDGA-treated α-synuclein was further analyzed to determine any structural effects of the interaction. Native nano-electrospray ionization mass spectra (ESI-MS), collected following 10 minute incubations of the protein-ligand mixtures, detected α-synuclein-NDGA complexes with masses matching the theoretical values of the unmodified protein and ligand[65]. This result indicates that the initial interaction between α-synuclein and NDGA does not require, nor produce, covalent modifications. Complexes were observed with higher relative intensities at low protein charge states, suggesting a preferential binding to the more compact α-synuclein conformations as previously observed with EGCG. However, levels of α-synuclein binding were much lower for NDGA than previously observed with EGCG[49] (Supplementary Fig. S7).

Ion-mobility mass spectrometry (IM-MS) allows for highly sensitive detection of changes in ion compactness, mass, and charge. Of note, this allows subtle changes in protein conformation e.g. caused by ligand binding to alternative sites, to be distinguished by compactness even at the same mass and charge of the ion. Further, native IM-MS can resolve multiple different binding events by their conformational effects without averaging across the species distribution. This capability has previously been employed to show the differential effects of EGCG and dopamine on α-synuclein compaction[49]. IM-MS revealed that α-synuclein compaction was unchanged by incubation with either NDGA or mNDGA (Figs 3A and S7). However, incubation with cyclized NDGA analogs cNDGA and cNDGA-5 led to α-synuclein assuming more compact conformations (Fig. 3B). α-Synuclein did non-covalently bind cyclized analogs up to 3 times, as seen by the corresponding mass shifts, and further, interaction with greater numbers of cyclized analogs correlated with increased compaction. This finding recapitulates similar results after α-synuclein incubation with EGCG[49] and indicates that interaction with cyclized NDGA causes structural compaction of α-synuclein.

NMR allowed for localization of NDGA's structural effects within the sequence of α-synuclein. After 24 hours incubation with 1:1 NDGA small variations in peak positions were observed in the N-terminus – Val 3, Phe 4, Met 5 – as well as His 50 (Fig. 3C). cNGDA and NDGA-1-treatment caused similar changes in the spectra (data not shown). Incubation with 3:1 NDGA resulted in more pronounced shifts in the same set of peaks (Supplementary Fig. S8).

These samples were followed for several days through NMR revealing a progressive variation of the peak shifts. After several days, additional changes occurred in the signals of several peaks. These peaks constitute a second set of signals, indicative of a subset of the α-synuclein population with altered conformation at these residues.

Förster resonance energy transfer (FRET) analysis of α-synuclein provided further insight into the regions altered by NDGA treatment. This examination employed several synthetic α-synuclein proteins with varying fluorescent label pairs spread throughout the amino acid sequence. Based on this methodology, FRET signal from an individual label pair can be converted to an intramolecular distance within α-synuclein and, when taken together, the multiple label pairs provided eight partially overlapped measurements covering the amino acid sequence from residue 9 to 136[66]. FRET revealed progressive changes in α-synuclein conformation when treated with EGCG, while NDGA, cNDGA, and mNDGA did not induce changes, even after 24 hours (Fig. 3D,E). As with IM-MS and NMR, NDGA-treatment was found to induce minimal changes in α-synuclein conformation, while EGCG showed more marked effects.

NDGA-treated and dialyzed α-synuclein examined by CD showed no changes in secondary structure further confirming the IM-MS and NMR analyses (Fig. 3F). Despite observing limited effects of NDGA on α-synuclein structure in aqueous solution, we also examined whether NDGA treatment alters the capacity of α-synuclein to interact with hydrophobic membranes. Membrane interaction induces the N-terminus of α-synuclein to assume an α-helical secondary structure[67–69]. Capacity for membrane interaction is necessary for α-synuclein's putative biological functions in the synapse[15,18,19]. Neither NDGA nor mNDGA treatment prevented α-synuclein from assuming α-helical conformations in the presence of SDS, just as occurred under control conditions (Fig. 3G). This indicates that the structural changes induced by NDGA do not restrict α-synuclein dynamic flexibility.

The ability of NDGA-treated α-synuclein to interact with membranes was further examined by fluorescence correlation spectroscopy (FCS), which has been used extensively to study the interaction of α-synuclein with lipid vesicles[16,70]. The binding of fluorescently labelled recombinant human α-synuclein to various concentrations of
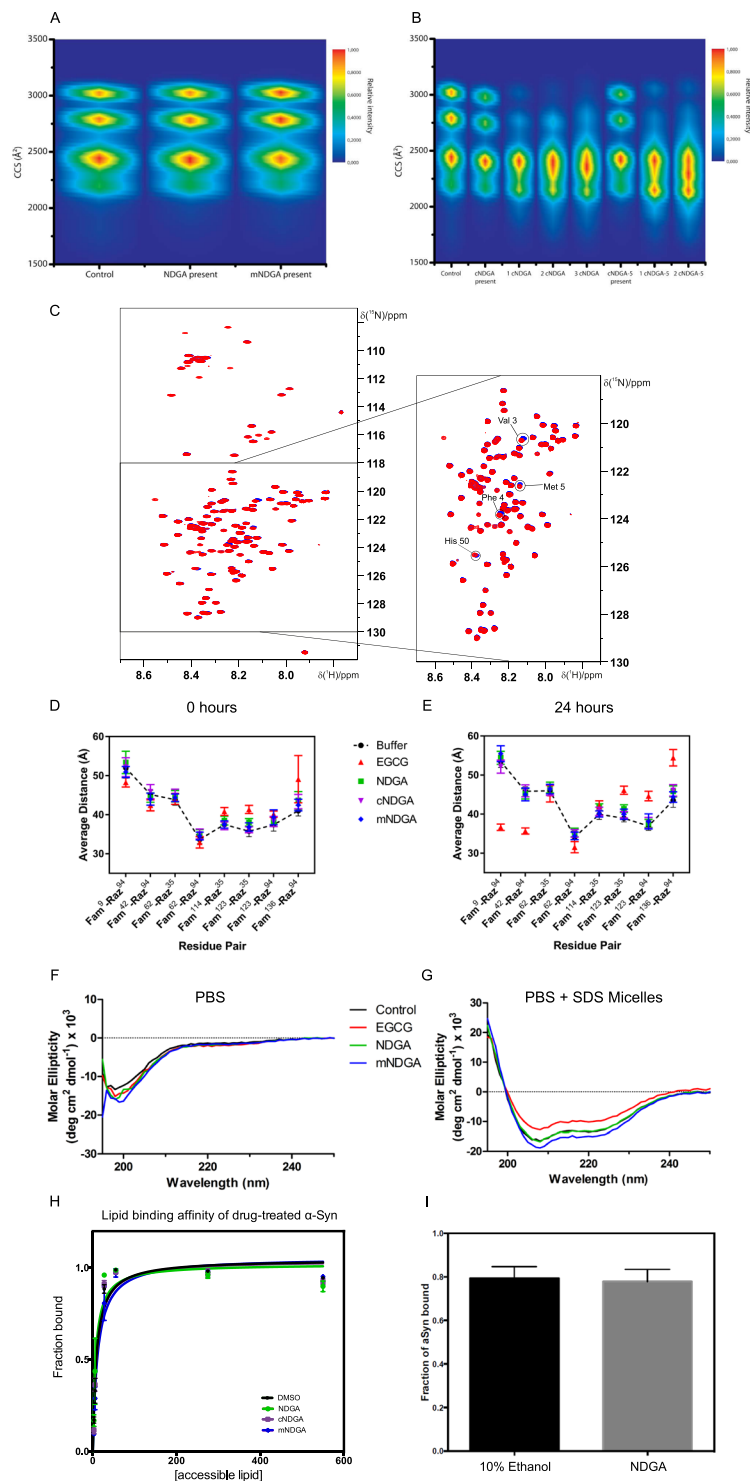
Figure 3. NDGA induces compaction of α-synuclein without preventing structural remodeling. (**A**) NDGA and mNDGA treatment did not alter α-synuclein collisional cross section measured by IM-MS. (**B**) cNDGA and cNDGA-5 induced α-synuclein compaction as measured by IM-MS. α-Synuclein was incubated with each molecule at 5:1 molar excess for 10 minutes before measurement. (**C**) NDGA treatment of α-synuclein did not induce extensive shifts in 2D NMR spectra. α-Synuclein was incubated 1:1 with NDGA for 24 hours before spectra were collected. NDGA-treated α-synuclein spectra (blue) was overlaid on solvent-treated α-synuclein (red). Fluorescein-maleimide (Fam) and tetramethylrhodamide azide (Raz) residue FRET measurement of (**D**) 0 hour and (**E**) 24 hour small molecule treatments showed progressive alteration of α-synuclein intramolecular distances by EGCG, but not NDGA, cNDGA, or mNDGA. The dashed line depicts 1 μM α-synuclein treated with buffer (1x PBS). Treatments were 5 μM NDGA (green square), EGCG (red triangle), cNDGA (purple inverted triangle), and mNDGA (blue diamond) (n = 3). (**F**) α-Synuclein secondary structure was not altered by pretreatment with NDGA. (**G**) NDGA pretreatment did not prevent α-synuclein assuming α-helical

secondary structure in the presence of SDS micelles. α-Synuclein was incubated 1:1 with NDGA, mNDGA, or solvent alone for 24 hours and then dialyzed against PBS for 24 hours. 40 mM SDS micelles were added 5 minutes before analysis. Secondary structure was quantified by circular dichroism. (**H**) NDGA treatment did not alter α-synuclein phospholipid binding. α-Synuclein was incubated 1:1 with NDGA analogs or solvent alone for 24 hours before fluorescence correlation spectroscopy in the presence of POPS:POPC vesicles at the indicated concentrations (n = 3). (**I**) Addition of NDGA did not displace fluorescently labeled α-synuclein from POPS:POPC vesicles (n = 3).

phospholipid vesicles (1:1, POPS:POPC) was compared following α-synuclein incubation for 24 hours with equimolar NDGA analogs or solvent alone. NDGA, cNDGA, and mNDGA did not alter phospholipid affinity as indicated by similar protein binding affinity at each vesicle concentration, regardless of treatment condition (Fig. 3H). FCS was also used to determine whether the addition of NDGA would displace α-synuclein already bound to phospholipid vesicles. The fraction of α-synuclein bound to vesicles was not altered by the addition of NDGA as compared to the addition of solvent alone (Fig. 3I). These data further indicate that NDGA treatment does not perturb α-synuclein interaction with hydrophobic membranes. Similarly, EGCG did not disrupt phospholipid or vesicle binding (Supplementary Fig. S9). Collectively, these results show that NDGA causes α-synuclein compaction across the sequence, without altering secondary structure or preventing structural remodeling involved in physiological function.

NGDA pretreated α-synuclein stably resists seeded aggregation. Based on the progressive remodeling of α-synuclein treated with NDGA, we analyzed the capacity of NDGA-pretreated α-synuclein to aggregate into fibrils. α-Synuclein was incubated with equimolar NDGA (as well as mNDGA and EGCG controls) for 24 hours before dialysis against excess PBS for 24 hours. This pretreated, dialyzed α-synuclein was then aggregated for 3 days before undergoing the solubility assay (as depicted in Supplementary Fig. S10). NDGA, but not mNDGA, pretreatment prevented α-synuclein aggregation (Fig. 4A). Further, pretreated, dialyzed α-synuclein was aggregated for 14 days to test the duration of the effect. Again, NDGA but not mNDGA pretreatment prevented α-synuclein aggregation (Supplementary Fig. S11). Consistent with previous data, we found that pretreatment with cyclizing (NDGA and NDGA-5) and cyclized (cNDGA and cNDGA-5) analogs prevented α-synuclein aggregation, while non-cyclizing analogs (NDGA-1 and SECO-1) did not (Supplementary Figs S12 and S13). We also examined the soluble species that remained following aggregation of pretreated, dialyzed α-synuclein using native state SEC and found that monomeric species are the majority product (Supplementary Fig. S14). Collectively, these findings demonstrate the NDGA pretreatment renders α-synuclein monomers resistant to aggregation. This effect was not observed with non-oxidizable and non-cyclizing analogs of NDGA.

Native-state size exclusion chromatographic (SEC) fractionation found that pretreatment with either NDGA or EGCG produces both monomeric and oligomeric α-synuclein species (Supplementary Fig. S6). We compared the aggregation of these two pools of α-synuclein species. α-Synuclein was incubated with equimolar concentrations of small molecules for 24 hours before oligomeric (≥51 Å) and monomeric (<51 Å) species were separated using native state SEC. The resulting fractions were then aggregated for 3 days before analysis by the solubility assay. Both monomers and oligomers produced upon NDGA and EGCG pretreatment resisted aggregation, while the majority of mNDGA treatment products formed insoluble aggregates (Fig. 4b). This indicates that while aggregation-resistant α-synuclein oligomers do form after NDGA treatment, resulting monomers are also aggregation-resistant, even in the absence of oligomers.

Next, we explored whether NDGA can prevent surface-catalyzed, secondary nucleation-dependent, aggregation of α-synuclein. Pretreated, dialyzed α-synuclein was aggregated for 3 days in the presence of α-synuclein fibril seeds (equal to 5% of total α-synuclein in solution) before aggregation was quantified by the solubility assay. Again, NDGA but not mNDGA pretreatment reduced α-synuclein aggregation (Fig. 4c). This demonstrates that NDGA-pretreated, dialyzed α-synuclein does not readily undergo secondary nucleation to form species capable of incorporating into existing fibrils.

Based on the observation that NDGA treatment prevents α-synuclein aggregation, we conducted a novel analysis to determine whether aggregation of untreated α-synuclein is altered by the presence of NDGA-pretreated α-synuclein. Pretreated, dialyzed α-synuclein was generated as before, then mixed with untreated α-synuclein at varying ratios and the resulting mixtures were aggregated for 3 days before undergoing the solubility assay (as depicted in Supplementary Fig. S15). The addition of pretreated, dialyzed α-synuclein caused a super-stoichiometric reduction in insoluble α-synuclein, demonstrating that aggregation of untreated α-synuclein was reduced (Fig. 4d). This effect is demonstrated in the 20% condition wherein NDGA pretreated, dialyzed α-synuclein represents only 20% of total α-synuclein, but insoluble α-synuclein is reduced from 88% in control conditions to 37%. These data, taken in totality, indicate that pretreatment with NDGA produces α-synuclein monomers that are not only resistant to aggregation, but reduce aggregation of α-synuclein never exposed to NDGA.

cNDGA reduces α-synuclein-driven dopamine neurodegeneration in animals. We examined the effect of NDGA and cNDGA treatment on α-synuclein-driven neurodegeneration (Fig. 5A,B). We employed a widely used, *Caenorhabditis elegans (C. elegans)* model in which expression of human wildtype α-synuclein in dopaminergic neurons leads to progressive neurodegeneration[71]. Animals were treated with varying doses of NDGA, or cNDGA from hatching until the day before scoring on either day 6 or day 8. Animals were considered to be undergoing neurodegeneration if any dopamine neurons were absent or dendritic processes showed signs of dysfunction such as blebbing[72]. Representative images of animals scored during this experiment show
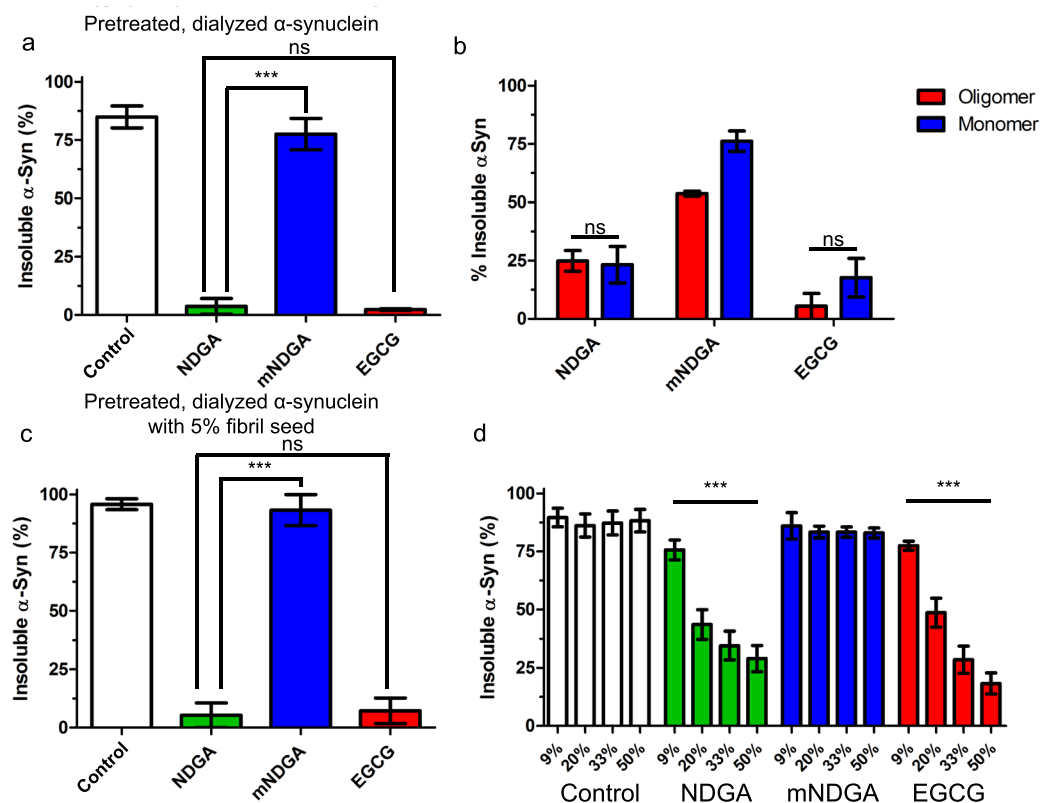
Figure 4. NDGA pretreatment prevents α-synuclein aggregation. (**a**) α-Synuclein did not aggregate after pretreatment with NDGA. α-Synuclein was incubated 1:1 with small molecules for 24 hours then dialyzed against PBS for 24 hours. After aggregation for 3 days, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry (n = 3, ***p < 0.001). (**b**) Both oligomeric and monomeric α-synuclein species induced by NDGA treatment resist aggregation. α-Synuclein was incubated 1:1 with small molecules for 24 hours then subjected to native state size exclusion chromatography. Oligomeric (≥51 Å) and Monomeric (<51 Å) α-synuclein fractions were collected and aggregated separately for 3 days. Soluble and insoluble species were separated and quantified as above (n = 3). (**c**) NDGA-treated α-synuclein resisted fibrillization in the presence of 5% fibril seed. α-Synuclein was incubated 1:1 with small molecules for 24 hours then dialyzed against excess PBS for 24 hours. Untreated α-synuclein fibrils equal to 5% total protein was added immediately before mixtures were aggregated for 3 days. Soluble and insoluble species were separated and quantified as above (n = 3, ***p < 0.001). (**d**) NDGA pretreated, dialyzed α-synuclein inhibited aggregation of untreated α-synuclein. α-Synuclein was incubated 1:1 with small molecules for 24 hours then dialyzed against excess PBS for 24 hours. Pretreated, dialyzed α-synuclein was then mixed with untreated monomeric α-synuclein at the indicated ratios. Mixtures were aggregated for 3 days. Soluble and insoluble species were separated and quantified as above (n = 3, ***p < 0.001).

the loss of fluorescent dopamine neuron processes and bodies in solvent and NDGA-treated animals (Fig. 5B). No differences were observed in NDGA or cNDGA treated animal after day 6 (data not shown). cNDGA at all concentrations (10, 50 and 100 μM) significantly reduced the number of animals undergoing dopamine neuron degeneration at day 8 (Fig. 5A). Similarly, EGCG (50 μM) protected against neuron loss (Supplementary Fig. S16). In this model, expression of α-synuclein was limited to just six neurons with each animal. As such, total α-synuclein levels were so low as to preclude biochemical analysis of the interaction between α-synuclein and NDGA or cNDGA. Likewise, we were unable to conduct pharmacokinetic analysis of our target neurons. Despite these limited reservations, these results demonstrate, for the first time, that cNDGA can reduce neurodegeneration caused by α-synuclein.

Discussion

Because of the engagement in several neurodegenerative diseases, α-synuclein aggregation remains a promising target for therapeutic development. Using NDGA and novel analogs, we uncovered that NDGA oxidation and cyclization was required for formation of quinone-modified monomeric α-synuclein. This interaction caused slight compaction of the α-synuclein molecule but did not induce any noticeable changes in its intrinsically
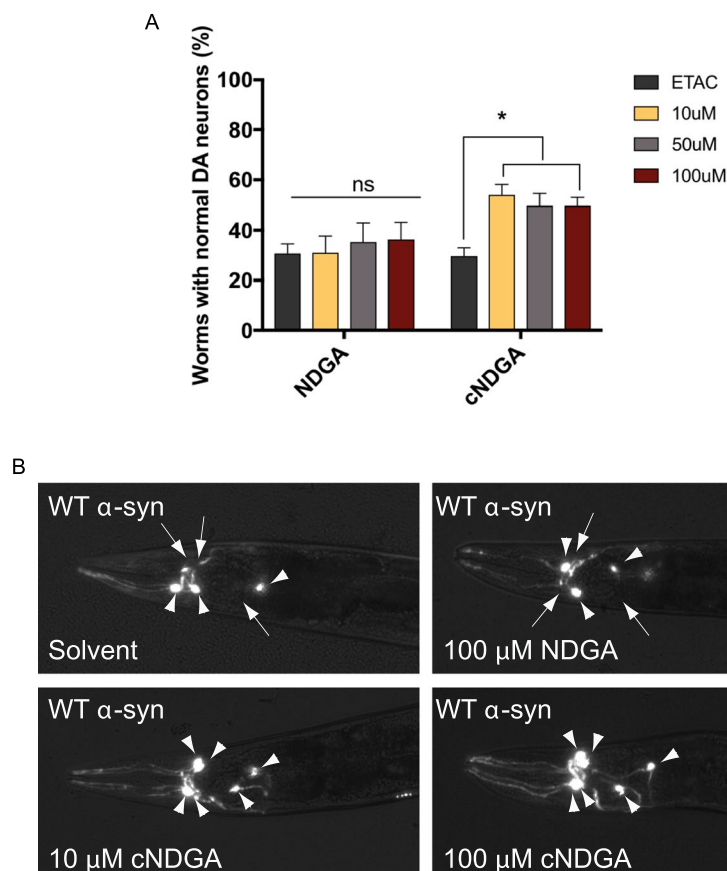
Figure 5. cNDGA reduces α-synuclein-driven neurodegeneration (**A**) cNDGA, but not NDGA reduces dopaminergic neurodegeneration in *C. elegans* expressing wildtype α-synuclein. Animals were exposed to each concentration of the drugs on days 0–3, 5, and 7 post-hatching. Animals were scored on day 8 post-hatching for dopaminergic neurodegeneration. The data are represented as mean ± SEM; one-way ANOVA with Tukey's *post hoc* test for multiple comparisons (n = 3; 30 animals per replicate; *$p < 0.05$; ns, not significant). (**B**) Representative images from day 8 post-hatching. Arrowheads indicate intact neurons while arrows indicate degenerating or missing neurons. Scale bar represents 10 μm.

disordered state. Furthermore, NDGA-treated α-synuclein retained its ability to undergo conformational changes required for interaction with the membranes, and NDGA did not dissociate protein-membrane complexes. However, interaction of α-synuclein with NDGA and its cyclizing and cyclized analogues efficiently prevented α-synuclein aggregation into fibrils. Conversely, non-cyclizing NDGA analogs had no effect on fibril formation, ensuring the aggregation resistant effects of NDGA and its analogs is not an artifact of reduced contact between α-synuclein molecules. Indeed, NDGA-treated monomers not only resisted aggregation, but prevented aggregation of naïve untreated α-synuclein. Treatment with cNDGA reduced neurodegeneration in an animal model of α-synuclein neurotoxicity. Taken together, these findings suggest that NDGA and related molecules that prevent α-synuclein aggregation without perturbing its native intrinsic disorder, or associations, may be promising targets for PD and related disorders characterized by α-synuclein aggregation.

Attempts to develop therapeutic strategies targeting α-synuclein have included decreasing levels by altering expression[73] and clearance[74,75], immunological targeting[76,77], and altering aggregation. Many aggregation inhibitors were identified through chemical library screens before being subjected to mechanistic analysis[32,40,41]. The neurotransmitter dopamine, and molecules that share structural and chemical similarities, were the first inhibitors identified and are among the most potent inhibitors of α-synuclein aggregation[32].

Identification of dopamine and related phenols as inhibitors of α-synuclein aggregation has led to many studies of phenolic compounds as potentially neuroprotective inhibitors. The most studied of these is EGCG, which is currently undergoing clinical trials in multiple system atrophy, a synucleinopathy characterized by microglial inclusions[78]. EGCG was found to be neuroprotective in animals treated with MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), which is a prodrug to the neurotoxin MPP$^+$ (1-methyl-4-phenylpyridinium) that induces α-synuclein-dependent dopaminergic neurodegeneration[79–81]. Adding EGCG during *in vitro* α-synuclein aggregation results in generation of less toxic aggregated products[39]. Given the structural differences between EGCG and NDGA, as well as the existing knowledge about NDGA oxidation and the availability of analogs, we hope

that studying NDGA as an α-synuclein aggregation inhibitor will provide new insight into the neuroprotective potential of inhibiting α-synuclein aggregation.

Our analysis revealed some substantial departures between the effects of NDGA and EGCG on α-synuclein. Konijnenberg *et al.* found, using IM-MS, that EGCG induces α-synuclein compaction[49], while analysis conducted by the same group for this study showed that cNDGA and cNDGA-5 induce α-synuclein compaction (Fig. 3A,B). NDGA, however, did not induce α-synuclein structural change. Interestingly, FRET found that EGCG causes substantial α-synuclein remodeling, while NDGA and cNDGA have no measurable effect on intramolecular distances (Fig. 3C). We also observed consistent differences in the patterns of SDS-stable multimers present after α-synuclein was treated with EGCG and NDGA (Figs 2B and S3, S6, S14). This trend of subtle, but measurable, differences between the effect of the two molecules leaves open the possibility of different mechanisms of their interaction with α-synuclein and differing biophysical properties of the resulting species.

Previous studies of EGCG have focused on the oligomers formed when α-synuclein is aggregated in the presence of EGCG[39,82,83]. In this study, we found that α-synuclein aggregation in the presence of EGCG and NDGA does produce oligomers, but that monomers are the predominant product, representing more than 75% of soluble protein (Supplementary Fig. S14). Further, we found that α-synuclein monomers and oligomers induced by pretreatment with EGCG and NDGA both resisted aggregation (Fig. 4b). The corresponding aggregation experiments were conducted after applying size-exclusion chromatography, which removed from solution any EGCG or NDGA not interacting with α-synuclein, suggesting a stable interaction or modification of α-synuclein itself. Most strikingly, NDGA-pretreated species inhibited aggregation of untreated α-synuclein, suggesting a dominant-negative effect on aggregation. This effect, not previously described for EGCG or any other exogenous phenolic inhibitors, may contribute to their neuroprotective effects.

There are two proposed models of amyloid-like fibril formation and elongation, although the precise mechanism remains unclear. In the first model monomers form oligomers and oligomers then stack together to form fibrils. In the second model of fibril formation, oligomers are elongated to fibrils with the addition of monomers. Under either paradigm, the data indicate that NDGA or EGCG modified oligomers do not progress to insoluble species. Possibly due to the increased compaction, monomeric α-synuclein treated with NDGA analogs, when incorporated into oligomers or fibrils may prevent further elongation. Essentially, the addition of a NDGA analog-modified α-synuclein monomer will "cap" the aggregation intermediate preventing the further addition of either monomer or oligomer. The stoichiometry would indicate that under either model the incorporation of aggregation resistant NDGA or EGCG modified monomer into oligomers or fibrils is sufficient to prevent further aggregation. The combination of a durable modification of α-synuclein and a dominant negative effect on aggregation are very desirable, given concerns over polyphenol bioavailability in the brain[84] and the hepatotoxic side effects seen with EGCG or NDGA[78,85].

Lipid-bound α-synuclein is resistant to aggregation[86], and appropriate membrane interactions are implicated in its physiological function[87]. As such, an ideal therapeutic targeting α-synuclein aggregation would allow normal lipid interaction. Interestingly, both CD and FCS showed that the membrane interactions of α-synuclein were not perturbed by NDGA-treatment (Fig. 3G–I). Preservation of α-synuclein membrane interaction may help explain the divergent effects of NDGA and dopamine on neuron viability. Indeed, α-synuclein oligomers induce membrane destabilization and dopamine-induced oligomers may contribute to neurodegeneration in synucleinopathies[36,37]. Further, the ability of α-synuclein to undergo native structural dynamics is necessary for formation of multimers implicated in physiological function[59,60] and aggregation resistance[88,89]. Future studies aimed at identifying neuroprotective α-synuclein aggregation inhibitors would be well served by incorporating methods to determine whether membrane interactions and structural dynamics are maintained, as is the case with NDGA.

NDGA pretreatment induced an NMR peak shift at His 50 (Fig. 3C). His 50 is located within the second region that assumes an α-helix during α-synuclein interaction with curved lipid membranes, falling in sequence just after the fourth lysine-rich repeat[67,69]. His 50 also mediates an oligomer-stabilizing interaction between α-synuclein and divalent metal ions, particularly Cu(II)[90,91]. Additionally, His 50 to Gln mutation causes autosomal dominant familial PD and accelerates α-synuclein aggregation *in vitro*[92–94]. This suggests that interaction with His 50 could alter α-synuclein aggregation kinetics. Further investigation of the interaction between NDGA and His 50 might provide insight into NDGA's inhibition of α-synuclein aggregation. Curiously, three other residues whose positions were perturbed by NDGA binding (Val 3, Phe 4, and Met 5) are located in the close proximity to the Asp 2, which is also engaged in Cu(II) binding[95].

The chemistry underpinning dopamine and related small molecule modulators of α-synuclein aggregation includes the capacity for auto-oxidation and the presence of vicinal hydroxyl (catechol) moieties[33,35]. In this study, we have expanded the previous observations and documented many new mechanistic insights for the biochemistry of catechol structures to inhibit the aggregation of α-synuclein. We found that NDGA-analogs require two pairs of vicinal hydroxyls to interact with α-synuclein. NDGA-1, which contains only one pair of vicinal hydroxyl groups, did not modify α-synuclein or inhibit aggregation. This indicates that NDGA cyclization, which is enabled by its two pairs of vicinal hydroxyls, is required for inhibition of α-synuclein aggregation[63]. Indeed, cNDGA and cNDGA-5, two cyclized analogs of NDGA, were found to inhibit α-synuclein aggregation. The finding that NDGA's cyclized oxidation products are responsible for its effects on α-synuclein aggregation presents the dual opportunities to study NDGA as a prodrug, with potential for conversion by oxidation-dependent cyclization to active cNDGA, and to examine novel molecules based on the cNDGA structure in hopes of identifying inhibitors of α-synuclein aggregation.

The importance of NDGA cyclization is reinforced by the finding that cNDGA, but not NDGA, treatment reduces neurodegeneration caused by expression of human α-synuclein in dopamine neurons in *C. elegans* (Fig. 5). There are many potential explanations for the divergent effects of the two molecules, including differences in stability, uptake, metabolism, and excretion. One possibility is that NDGA does not form cNDGA under these conditions. Oxidation of NDGA likely occurs less frequently in the reducing environment of the cell[96],

and yield of cNDGA may be lower than previously observed. While the cyclization of NDGA has been precisely described under controlled conditions in buffer[63], it has never been examined in a complex milieu or within a living cell, where other chemistry may occur.

It could be of interest to compare the molecular mechanism of action of NDGA on α-synuclein with the effects of different small molecules on functionality of other intrinsically disordered proteins. One of the best studied examples of such systems is given by small molecule-driven inhibition of heterodimerization of the transcription factor c-Myc with its partner, Max, via a basic-helix-loop-helix-leucine zipper (bHLHZip) domain present in both proteins. Importantly, both proteins in their unbound forms are disordered, but undergo mutual coupled binding and folding when their zipper domains interact to form a helical coiled coil[97–99]. High throughput screening uncovered several specific inhibitors that were able to bind to one of three discrete sites (residues 366–375, 375–385, and 402–409) within the 85-residue bHLHZip domain of the monomeric c-Myc[100,101]. Solution NMR analysis revealed that interaction of said small molecules with c-Myc resulted in its local misfolding, thereby generating conformations incompatible with the heterodimerization of this protein with Max[97–99]. Clearly, the mechanism of NDGA action, where a small molecule stabilizes specific conformations of intrinsically disordered α-synuclein, is principally different from the inhibitory activity of small molecules inducing c-Myc misfolding. Therefore, further study is warranted to determine whether the interaction between NDGA and α-synuclein represents a new mode of modulatory action of small molecules on intrinsically disordered proteins. Moreover, in-cell NMR data indicated that molecular crowding agents present in the cytosol induce compaction of α-synuclein and shielding of the aggregation prone region (hydrophobic core between residues 61–95) preventing oligomerization and aggregation[102]. Although ensemble-based techniques (FRET or NMR) are not able to visualize and inform on the structures of these compact states, we speculate that the compact states observed by NDGA analogs and EGCG are similar to those observed with in-cell NMR.

Collectively, these findings change our understanding of the mechanisms of phenolic α-synuclein aggregation inhibitors. We demonstrate, for the first time, that inhibition of α-synuclein aggregation by NDGA requires oxidation-dependent cyclization, reframing NDGA as a prodrug. Additionally, we show that both NDGA and EGCG stably modify α-synuclein monomers, rendering them aggregation incompetent. In the case of NDGA, this is achieved without alteration of α-synuclein secondary structure in solution, nor perturbing membrane interactions. This combination of attributes has not been previously observed in other inhibitors of α-synuclein aggregation. Further examination is certainly warranted, including a full characterization of efficacy and critically toxicity, to determine whether NDGA analogs might provide the basis for novel neuroprotective therapies for PD and related synucleinopathies.

Methods

***In vitro* α-synuclein aggregation analyses.**   Recombinant human wild type α-synuclein was expressed and purified from *E. coli* as previously described[103]. Purified α-synuclein was used at 2 mg/ml (138 μM) with or without small molecules at various molar ratios in PBS with 1% DMSO. Mixtures were aggregated for 3 or 7 days at 37 °C shaking at 1400 rpm. The longer 7 day incubation was performed to ensure the treatment effects remained consistent even under prolonged aggregation. Alternatively, purified α-synuclein was incubated for 24 hours with or with equimolar small molecules in PBS with 1% DMSO. Resulting mixtures were then dialyzed against excess PBS for 24 hours at 4 °C (using mini-dialysis tubes, Thermo Fisher 69572). Mixtures were aggregated for 3 or 14 days shaking at 1400 rpm. Amyloid fibril formation was monitored by thioflavin T. Briefly, 4 μM protein was incubated with 25 μM ThT (Sigma Aldrich) dissolved in PBS and read at Ex of 450 and Em 482 using a fluorescent plate reader (Molecular Devices).

α-Synuclein solubility analysis and densitometric quantification.   Sedimentation was performed by centrifugation at 21,000 g for 15 minutes. Supernatants were removed, and pellets were resuspended in an equal volume. Pellets and supernatants were boiled in SDS sample buffer for 3 minutes at 95 °C. α-Synuclein species were resolved by SDS-PAGE in 12% Bis/Tris gels and stained in-gel with colloidal blue (Invitrogen LC6025). Stained gels were imaged at 700 nm using LI-COR Odyssey Infrared Imaging System and densitometric quantification was conducted using LI-COR Odyssey software suite 3.0.

Near infrared fluorescence.   SDS-PAGE gels were imaged as previously described[64]. Briefly, gels were imaged immediately after electrophoresis at 700 nm with intensity 10 on LI-COR Odyssey Infrared Imaging System.

Size-exclusion chromatography.   For size exclusion chromatography 200 μL recombinant α-synuclein (2 mg/mL) was injected onto a Superdex 200 10/300 GL (GE Healthcare) connected to an Agilent 1100 series HPLC system and fraction collector controlled by ChemStation software version 1.04 (Agilent). Mobile phase consisted of 25 mM HEPES and 150 mM NaCl, pH 7.25 with a 0.3 mL/min flow rate. Fractions corresponding to 140-100, 100-85, 85-75, 75-68, 68-61.5, 61.5-56, 56-51, 51-47, 47-43, 43-39, 39, 39-36, 36-32, 32-29, 29-26 Å were combined and concentrated using 3 kDa NMWL ultracel microcon filters (Millipore UFC5003). SEC column elution time was calibrated to Å size using globular protein standards (GE Healthcare).

Circular dichroism.   Circular dichroism spectra were obtained on a Jasco-810 spectropolarimeter maintained in the Children's Hospital of Philadelphia Protein Core facility. Protein in PBS was diluted to 20 μM in 0.05 M KH_2PO_4 pH 7.8. Spectra were corrected for baseline measurement of an equivalent volume of PBS diluted in KH_2PO_4 buffer. Spectra were collected with a scanning width of 1 nm, at 5 nm/min with 2 accumulations per run.

Transmission electron microscopy. Transmission electron microscopy was conducted on a JEOL1010 maintained in the University of Pennsylvania Electron Microscopy Resource Laboratory. Protein solutions were diluted to 0.5 mg/mL (34.5 μM) before mounting to carbon-coated 300 grids and negative staining with 2% uranium acetate. Images were collected with HV = 80.0 kV, Magnification 50,000x. Contrast was automatically adjusted during capture for each image, and not altered after capture. 1 μm scale bars were included for each image presented.

Nuclear magnetic resonance spectroscopy. To provide atomic resolution information on α-synuclein, 2D ^{1}H-^{15}N NMR spectra were acquired through a series of band-selective excitation short-transient transverse relaxation optimized spectroscopy heteronuclear single quantum coherence (BEST-TROSY HSQC) runs[104], a fast and common experiment used to evaluate if there is an effect of the different molecules on some amino acids, each of them represented by a H^N-N correlation (a cross-peak in the 2D spectrum).

All the NMR experiments were recorded on a Bruker Avance III spectrometer operating at 900 MHz ^{1}H frequency (21.14 T) equipped with a cryogenically cooled probehead for triple resonance experiments (TCI). The pulses involved in this pulse sequence were the standard ones used to investigate bio-molecules[105]. The ^{15}N hard pulses were applied at the center of the region (117.02 ppm) and the ^{1}H band selective pulses were applied at the center of the amide proton region at 8.70 ppm. A spectral width of 10822.511 Hz in the direct ^{1}H dimension and 3647.991 Hz in the indirect ^{15}N dimension was used. All the spectra were recorded with 8 scans per increment, 4096 and 2048 points in the direct dimension and in the indirect dimension respectively and a recycle delay of 500 ms.

The experiments were recorded at 288.0 K on a 165 μM uniformly ^{15}N labeled α-synuclein sample, prepared as previously described[106], in buffer (10 mM KPi, 140 mM NaCl, 0.25 mM EDTA + Roche protease inhibitors, pH 7.5). 5% D$_2$O was added for the lock signal. The different drugs (described above) were added as DMSO solutions at 1:1 stoichiometric ratio and left at room temperature 24 hours to incubate before NMR measurements. The same experiments were acquired on a sample containing 3 equivalents of NDGA. Care was taken to minimize the volume of the DMSO solution added and to keep it constant for all the samples investigated. All the samples were analyzed again in the following days to evaluate thermodynamic and kinetic effects.

Native nano-electrospray ionization mass spectrometry. ESI-MS experiments were conducted as previously described[49]. The only methodological departure was the inclusion of DMSO to solubilize the ligands. Briefly, MS spectra were collected after 10 minute incubation of protein-ligand mixtures in pH 7.4 10 mM ammonium acetate. Spectra were collected using a hybrid quadrupole-time-of-flight mass spectrometer (QSTAR-Elite, Biosystems, Foster City, CA) equipped with a nano-ESI sample source.

Native Ion mobility mass spectrometry. Native Ion mobility-mass spectrometry (IM-MS) experiments were conducted as previously described[49]. The only methodological departure was the inclusion of DMSO to solubilize the ligands. Briefly, MS spectra were collected after 10 minute incubation of protein-ligand mixtures in pH 7.4 10 mM ammonium acetate. α-Synuclein was present in the mixtures at 20 μM, ligand at 100 μM. IM-MS was performed on a Synapt G2 HDMS (Waters, Manchester, U.K.) using nano-ESI with homemade gold-coated borosilicate capillaries.

Fluorescence correlation spectroscopy. *Expression of fluorescently-labeled α-synuclein.* α-Synuclein protein labeled at residue S9C with Alexa Fluor 488 was produced via recombinant protein production. Plasmid containing α-synuclein S9C fused to a polyhistidine-tagged GyrA intein from Mycobacterium xenopi (Mxe) was transformed into BL21 DE3 competent cells by heat shocking at 42 °C. Single colonies grown on Ampicillin (Amp) plates were picked to inoculate primary cultures in LB supplemented with 1 μg/mL Amp. Primary cultures were combined into secondary cultures, which were grown at 37 °C in a shaker-incubator until optical density (OD) reached ~0.6. Expression of the gene of interest was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were then grown in the shaker-incubator at 18 °C overnight. After centrifugation (5000 rpm, 20 min, 4 °C), cell pellets were re-suspended in re-suspension buffer (20 mM Tris, 1 mM PMSF, 1 Roche protease inhibitor tablet, pH 8.3) and sonicated in a cup in an ice bath (5 min, 1 s ON, 1 s OFF). The resulting lysate was centrifuged (14,000 rpm, 25 min, 4 °C), and supernatant containing the protein of interest (POI) was purified over Ni-NTA affinity column. Intein cleavage was carried out by incubation with 200 mM β-mercaptoethanol (βME) on a rotisserie over night at room temperature. Cleaved POI was dialyzed into 20 mM Tris pH 8 before purification over a second Ni-NTA column to remove the free intein from the sample. Flow-through containing the POI was kept, dialyzed into 20 mM Tris pH 8, and spin concentrated. Labeling with Alexa Fluor 488 (AF488) maleimide was done by adding 4 equivalents (eq.) of fluorophore dissolved in DMSO. The reaction tube was wrapped in aluminum foil and incubated at 37 °C for ~6 hrs. Formation of the product α-synuclein C9-AF488 was checked by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and polyacrylamide gel electrophoresis (SDS-PAGE). Labeled protein was purified by fast-protein liquid chromatography (FPLC) using a Hi-Trap Q 5 mL column and by reverse-phase high-performance liquid chromatography (RP-HPLC) using a C18 preparatory column. Purified protein was spin concentrated into buffer (20 mM Tris, 100 mM NaCl), and aliquots were stored at −80 °C.

Treating α-synuclein with drug molecules. The concentration of α-synuclein aliquots was determined by UV-Vis at 488 nm. The drug molecules NDGA, mNDGA, and cNDGA were dissolved in DMSO and diluted 10-fold into buffer (20 mM Tris, 100 mM NaCl). A control of 10% DMSO in buffer was also made. Each aliquot of α-synuclein was respectively treated with 1eq. of NDGA, mNDGA, and cNDGA, and DMSO-buffer. Treatment samples were incubated at room temperature for 8 hours, then dialyzed against 10 mM Tris, 100 mM NaCl overnight.

Preparation of lipid vesicles.   Synthetic lipid vesicles were prepared for use in binding experiments. A mixture in 50:50 molar ratio of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine (POPS) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) were drawn from chloroform stock and dried under nitrogen gas to form a film inside a glass vial. Films were desiccated under vacuum and re-hydrated in 3-(N-morpholino) propanesulfonic acid (MOPS) buffer. 10 Freeze-thaw cycles consisting of cooling in liquid nitrogen for 40 s and warming in a 60 °C water bath for 2 min were performed to aid the formation of uniformly sized vesicles. Using syringes fitted onto an extruder, vesicles were pushed 31 times through 50 nm pore membranes. Vesicles were determined by Dynamic Light Scattering to be monodisperse and distributed uniformly around 80 nm in diameter, consistent across different concentrations of all samples. All lipid vesicles were prepared fresh and used within 48 hours of extrusion.

Fluorescence correlation spectroscopy.   Eight-well chambered coverglasses (Nunc, Rochester, NY) were prepared by plasma cleaning followed by incubation over night with polylysine-conjugated polyethylene glycol (PEG-PLL), prepared using a modified Pierce PEGylation protocol (Pierce, Rockford, IL). PEG-PLL coated Chambers were rinsed with and stored in Milli-Q water until use. FCS measurements were done on a lab-built instrument based on an Olympus IX71 microscope with a continuous emission 488 nm DPSS 50 mW laser (Spectra-Physics, Santa Clara, CA). All measurements were done at 20 °C. The laser power entering the microscope was adjusted to 4.5 μW. Fluorescence emission collected through the objective was separated from the excitation signal through a Z488rdc long pass dichroic and an HQ600/200 m bandpass filter (Chroma, Bellows Falls, VT). Emission signal was focused through a 50 μm optical fiber. Signal was amplified by an avalanche photodiode (Perkin Elmer, Waltham, MA) coupled to the fiber. A digital autocorrelator (Flex03Q-12, correlator.com, Bridgewater, NJ) was used to collect 30 autocorrelation curves of 30 seconds for each measurement. Fitting was done using MATLAB (The MathWorks, Natick, MA).

Binding assay of drug-treated α-synuclein to lipid vesicles.   FCS was used to examine the binding affinity of drug-treated α-synuclein to lipid vesicles. Each drug-treated α-synuclein labeled with AF488 was examined in the presence of varying concentrations of lipid vesicles, and each autocorrelation curve was fit to a 2-component equation to extract the fraction of bound α-syn at each concentration. From these data, a binding curve was generated and fit to an exponential to determine the dissociation constant Kd, i.e. the concentration at which half of the protein is bound. In fitting the autocorrelation curves for α-syn in the presence of lipid vesicles, the diffusion time (τ) of bound and unbound α-syn were respectively fixed to experimentally determined values. The diffusion time of unbound protein, $\tau_{\alpha S}$, was determined by measurements of the protein in buffer without lipids. Since bound protein diffuses with the vesicles to which they are bound, the diffusion time of the vesicles, $\tau_{vesicle}$, was determined by measurements of BODIPY-labeled 50:50 POPS/POPC. The diffusion time of our unlabeled vesicles were also deduced from a calibration curve generated from the diffusion times of commercial fluorescent bead standards of 50 nm and 100 nm in diameter.

To generate a vesicle-binding curve, FCS was performed on α-syn C9-AF488 in the presence of varying concentrations (0.005 mM to 1 mM lipid) of 50:50 POPS/POPC. The accessible surface lipid concentration was calculated based on the characteristic bilayer thickness of POPS and POPC. The fraction of bound protein was extracted from the fit to each autocorrelation curve by fixing $\tau_{\alpha S}$ and globally fitting $\tau_{vesicle}$ to the same value across all concentrations.

$$G(\tau) = \frac{1}{N}\left(F_{F*}\frac{1}{1+\frac{\tau}{\tau_{\alpha S}}}*\left(\frac{1}{1+\frac{s^2\tau}{\tau_{\alpha S}}}\right)^{1/2} + (1-F_F)*\frac{1}{1+\frac{\tau}{\tau_{vesicle}}}*\left(\frac{1}{1+\frac{s^2\tau}{\tau_{vesicle}}}\right)^{1/2}\right)$$

The fraction bound was plotted against accessible lipid concentration and fit to an exponential, from which the Kd was determined.

$$F_B = \frac{B_{max}x}{K_{d,app}+x}$$

Assay for effect of NDGA on lipid-bound α-synuclein.   Untreated α-synuclein C9-AF488 (final concentration 20 μM) was added to 250 μL of 0.05 mM POPS/PC in a microscope chamber well. The sample was incubated for 15 minutes to ensure maximum binding interactions. NDGA stock was made by dissolving the drug in ethanol and diluting 10-fold into buffer (20 mM Tris, 100 mM NaCl). NDGA was added at 20 μM final concentration to the vesicle-bound α-synuclein sample, and fraction of protein bound was determined using the fitting described above. A control treatment was made of 10% ethanol in buffer, and fraction of protein bound was determined in the same manner.

Förster resonance energy transfer anaylsis.   Small molecule stock solutions containing either NDGA, EGCG, cNDGA or mNDGA were prepared in DMSO. Labeled proteins constructs, including new constructs not previously described, were prepared using methods previously described[66]. Labeled positions not previously reported were confirmed by mass spectrometry of both the full-length protein and fragments from trypsin digestion using MALDI. The concentration of each protein stock was assessed by UV-Vis absorbance of the attached fluorescein-maleimide (Fam) ($\varepsilon_{494} = 68{,}000\ M^{-1}\ cm^{-1}$) and/or tetramethylrhodamide azide (Raz)

($\varepsilon_{555} = 87,000\,M^{-1}\,cm^{-1}$) dyes. Fluorescence assays were performed in a black, clear bottom half-diameter Greiner 96-well plate and measurements were taken on a Tecan M1000 plate reader. Each well of the plate contained a single protein sample, containing either one or both fluorescent labels, and one of the small molecules of interest. Protein samples were also measured in the absence of small molecules as a control. Each sample was prepared in 1X PBS with a final concentration of $1\,\mu M$ protein. Samples containing NDGA, EGCG, cNDGA and mNDGA were prepared with a 5-fold excess of small molecules relative to the protein concentration. Small molecule stocks were prepared just prior to performing the assay. Each assay was performed by diluting the protein in buffer in each well and adding the small molecule solution, using a multichannel pipette, just prior to measuring the fluorescence. Each sample was excited at 494 nm and the emission spectrum was measured from 502–700 nm with a 1 nm step size. Each step consisted of 50 flashes at a frequency of 400 Hz with a 20 μs integration time and the gain was set to 135. The Z-position of the plate was optimized prior to the first measurement and was maintained at 21728 μm for all measurements. All measurements were performed at room temperature. After measuring the emission spectrum from each well immediately after the introduction of small molecule, the plates were sealed with parafilm and covered in aluminum foil and left at room temperature for 24 hrs. After 24 hrs the fluorescence of each sample was measured again as detailed above. The FRET between Fam and Raz probes attached to a single protein was assessed by computing the fluorescence quenching of the donor Fam fluorophore induced by the presence of the Raz acceptor as previously described[66]. FRET efficiencies were converted to average interresidue distances using the Förster equation and a gaussian chain polymer model.

Analysis of dopaminergic neurodegeneration in *C. elegans*.   *Generation of transgenic nematodes.*   Transgenic *C. elegans* lines were generated by microinjection using previously described methods[107]. The strain UA294 (baEx175 *a,b,c* [P*dat-1*::WT a-syn, P*unc-54*::tdTomato]; *vtIs7*[P*dat-1*::GFP]) was generated by injecting with a solution of 50 ng/μL plasmid with either P*dat-1*::WT a-syn with a phenotypic marker (P*unc-54*::tdTomato, 50 ng/μL, for body wall muscle expression). Three distinct stable lines (*a, b, c*) were generated.

Analysis of dopaminergic neurodegeneration and NDGA analog treatments in C. elegans.   NDGA and cNDGA were used at final concentrations of 10 μM, 50 μM, or 100 μM in ethyl acetate (EtAc). NDGA, cNDGA, or solvent (EtAc) was added to the surface of bacterial lawn on the nematode growth medium (NGM) Petri plates (60 mm diameter) at the corresponding final concentration and allowed to dry under the hood. For EGCG, 50 μM was added into the NGM agar plates. Three resulting independent transgenic lines (*a, b, c*) were synchronized, and exposed to corresponding concentration of NDGA, cNDGA, or solvent (EtAc) from hatching to day 3, then were transferred and maintained on NGM plates until the day of analysis. Animals on NGM plates were treated with additional NDGA, cNDGA, or solvent on days 5 and 7. For EGCG treatment, worms were transferred to fresh NGM plates daily and transferred onto fresh EGCG containing plates on days 0–3 and 5 post-hatching. For dopaminergic neurodegeneration analyses, the transgenic animals were scored as described previously[72]. Briefly, on the day of analysis, the six anterior dopaminergic neurons (four CEP (cephalic) and two ADE (anterior deirid)) were examined in 30 randomly selected nematodes with the marker transgene (tdTomato) in the body wall muscle cells. Worms were considered normal when all six anterior neurons were present without any signs of degeneration, such as broken dendritic process, cell body loss, dendritic blebbing, or a missing neuron. In total, at least 90 adult worms were analyzed for each treatment condition, at least 30 from each independent transgenic line. An average of total percentage of worms with normal neurons was reported in the study.

Reagents.   Nordihydroguaiaretic acid, NDGA 74540 Aldrich; (−)- Epigallocatechin gallate, EGCG E4143 Sigma or 4524 Tocris; Teprameprocol/mNDGA, T3455 Sigma; Catalase, C9322 Sigma; NDGA analogs NDGA-1, SECO-1, NDGA-5, cNDGA, and cNDGA-5 were generated, purified, and validated by Krol *et al.* at the University of Saskatchewan as previously described[63].

Statistical analysis.   Statistical analysis was conducted using Prism 5 software (GraphPad). Unpaired students *t* test was used for comparisons between two groups. Comparisons involving multiple groups, such as molecule and dose comparisons, used two-way ANOVA with Tukey's *post hoc* test for multiple comparisons. In all cases $p < 0.05$ was considered statically significant.

Data Availability
All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

References
1. Spillantini, M. G. *et al.* alpha-Synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
2. Baba, M. *et al.* Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am. J. Pathol.* **152**, 879–84 (1998).
3. Polymeropoulos, M. H. *et al.* Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).
4. Krüger, R. *et al.* Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.* **18**, 106–108 (1998).
5. Chartier-Harlin, M. C. *et al.* α-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169 (2004).
6. Singleton, A. B. *et al.* α-Synuclein Locus Triplication Causes Parkinson's Disease. *Science* **302**, 841 (2003).
7. Wakabayashi, K., Yoshimoto, M., Tsuji, S. & Takahashi, H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. *Neurosci. Lett.* **249**, 180–2 (1998).
8. Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc. Natl. Acad. Sci. USA* **95**, 6469–73 (1998).
9. Gai, W. P., Power, J. H., Blumbergs, P. C. & Blessing, W. W. Multiple-system atrophy: a new alpha-synuclein disease? *Lancet (London, England)* **352**, 547–8 (1998).

10. Trojanowski, J. Q., Goedert, M., Iwatsubo, T. & Lee, V. M. Fatal attractions: abnormal protein aggregation and neuron death in Parkinson's disease and Lewy body dementia. *Cell Death Differ.* **5**, 832–7 (1998).

11. Takeda, A. *et al.* Abnormal accumulation of NACP/α-synuclein in neurodegenerative disorders. *Am. J. Pathol.* **152**, 367–372 (1998).

12. Lücking, C. B. & Brice, A. Alpha-synuclein and Parkinson's disease. *Cell. Mol. life Sci.* **57**, 1894–908 (2000).

13. Arawaka, S., Saito, Y., Murayama, S. & Mori, H. Lewy body in neurodegeneration with brain iron accumulation type 1 is immunoreactive for alpha-synuclein. *Neurology* **51**, 887–9 (1998).

14. Wakabayashi, K., Matsumoto, K., Takayama, K., Yoshimoto, M. & Takahashi, H. NACP, a presynaptic protein, immunoreactivity in Lewy bodies in Parkinson's disease. *Neurosci. Lett.* **239**, 45–8 (1997).

15. Abeliovich, A. *et al.* Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron* **25**, 239–52 (2000).

16. Middleton, E. R. & Rhoades, E. Effects of curvature and composition on α-synuclein binding to lipid vesicles. *Biophys. J.* **99**, 2279–2288 (2010).

17. Larsen, K. E. *et al.* Alpha-synuclein overexpression in PC12 and chromaffin cells impairs catecholamine release by interfering with a late step in exocytosis. *J. Neurosci.* **26**, 11915–22 (2006).

18. Nemani, V. M. *et al.* Increased Expression of α-Synuclein Reduces Neurotransmitter Release by Inhibiting Synaptic Vesicle Reclustering after Endocytosis. *Neuron* **65**, 66–79 (2010).

19. Murphy, D. D., Rueter, S. M., Trojanowski, J. Q. & Lee, V. M. Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons. *J. Neurosci.* **20**, 3214–3220 (2000).

20. Greten-Harrison, B. *et al.* α-Synuclein triple knockout mice reveal age-dependent neuronal dysfunction. *Proc. Natl. Acad. Sci.* **107**, 19573–19578 (2010).

21. Scott, D. & Roy, S. α-Synuclein inhibits intersynaptic vesicle mobility and maintains recycling-pool homeostasis. *J. Neurosci.* **32**, 10129–35 (2012).

22. Giasson, B. I. *et al.* Neuronal α-synucleinopathy with severe movement disorder in mice expressing A53T human α-synuclein. *Neuron* **34**, 521–533 (2002).

23. Yang, W. *et al.* Mutant Alpha-Synuclein Causes Age-Dependent Neuropathology in Monkey Brain. *J. Neurosci.* **35**, 8345–8358 (2015).

24. Recasens, A. *et al.* Lewy body extracts from Parkinson disease brains trigger α-synuclein pathology and neurodegeneration in mice and monkeys. *Ann. Neurol.* **75**, 351–62 (2014).

25. Luk, K. C. *et al.* Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949–953 (2012).

26. Sacino, A. N. *et al.* Brain Injection of α-Synuclein Induces Multiple Proteinopathies, Gliosis, and a Neuronal Injury Marker. *J. Neurosci.* **34**, 12368–12378 (2014).

27. Shimozawa, A. *et al.* Propagation of pathological α-synuclein in marmoset brain. *Acta Neuropathol. Commun.* **5**, 12 (2017).

28. Rey, N. L. *et al.* Spread of aggregates after olfactory bulb injection of α-synuclein fibrils is associated with early neuronal loss and is reduced long term. *Acta Neuropathol.* **135**, 65–83 (2018).

29. Li, J. Y. *et al.* Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nat. Med.* **14**, 501–503 (2008).

30. Kowall, N. W. *et al.* MPTP induces alpha-synuclein aggregation in the substantia nigra of baboons. *Neuroreport* **11**, 211–213 (2000).

31. Betarbet, R. *et al.* Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat. Neurosci.* **3**, 1301–1306 (2000).

32. Conway, K. A., Rochet, J. C., Bieganski, R. M. & Lansbury, J. Kinetic stabilization of the α-synuclein protofibril by a dopamine-α-synuclein adduct. *Science* **294**, 1346–1349 (2001).

33. Norris, E. H. *et al.* Reversible inhibition of α-synuclein fibrillization by dopaminochrome-mediated conformational alterations. *J. Biol. Chem.* **280**, 21212–21219 (2005).

34. Mazzulli, J. R. *et al.* Cytosolic Catechols Inhibit α-Synuclein Aggregation and Facilitate the Formation of Intracellular Soluble Oligomeric Intermediates. *J. Neurosci.* **26**, 10068–10078 (2006).

35. Meng, X., Munishkina, L. A., Fink, A. L. & Uversky, V. N. Molecular mechanisms underlying the flavonoid-induced inhibition of α-synuclein fibrillation. *Biochemistry* **48**, 8206–8224 (2009).

36. Mor, D. E. *et al.* Dopamine induces soluble α-synuclein oligomers and nigrostriatal degeneration. *Nat. Neurosci.* **20**, 1560–1568 (2017).

37. Danzer, K. M. *et al.* Different Species of α-Synuclein Oligomers Induce Calcium Influx and Seeding. *J. Neurosci.* **27**, 9220–9232 (2007).

38. Burbulla, L. F. *et al.* Dopamine oxidation mediates mitochondrial and lysosomal dysfunction in Parkinson's disease. *Science* **357**, 1255–1261 (2017).

39. Ehrnhoefer, D. E. *et al.* EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat. Struct. Mol. Biol.* **15**, 558–566 (2008).

40. Ono, K. & Yamada, M. Antioxidant compounds have potent anti-fibrillogenic and fibril-destabilizing effects for alpha-synuclein fibrils *in vitro*. *J. Neurochem.* **97**, 105–15 (2006).

41. Masuda, M. *et al.* Small molecule inhibitors of α-synuclein filament assembly. *Biochemistry* **45**, 6085–6094 (2006).

42. Meng, X., Munishkina, L. A., Fink, A. L. & Uversky, V. N. Effects of Various Flavonoids on the α-Synuclein Fibrillation Process. *Parkinsons. Dis.* **2010**, 1–16 (2010).

43. Fernandes, L. *et al.* An ortho-Iminoquinone Compound Reacts with Lysine Inhibiting Aggregation while Remodeling Mature Amyloid Fibrils. *ACS Chem. Neurosci.* acschemneuro.7b00017, https://doi.org/10.1021/acschemneuro.7b00017 (2017).

44. Di Giovanni, S. *et al.* Entacapone and tolcapone, two catechol O-methyltransferase inhibitors, block fibril formation of α-synuclein and β-amyloid and protect against amyloid-induced toxicity. *J. Biol. Chem.* **285**, 14941–14954 (2010).

45. Lorenzen, N. *et al.* How epigallocatechin gallate can inhibit α-synuclein oligomer toxicity *in vitro*. *J. Biol. Chem.* **289**, 21299–21310 (2014).

46. Wagner, J. *et al.* Anle138b: A novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease. *Acta Neuropathol.* **125**, 795–813 (2013).

47. Deeg, A. A. *et al.* Anle138b and related compounds are aggregation specific fluorescence markers and reveal high affinity binding to α-synuclein aggregates. *Biochim. Biophys. Acta* **1850**, 1884–90 (2015).

48. Li, J., Zhu, M., Rajamani, S., Uversky, V. N. & Fink, A. L. Rifampicin inhibits alpha-synuclein fibrillation and disaggregates fibrils. *Chem. Biol.* **11**, 1513–21 (2004).

49. Konijnenberg, A. *et al.* Opposite Structural Effects of Epigallocatechin-3-gallate and Dopamine Binding to α-Synuclein. *Anal. Chem.* **88**, 8468–8475 (2016).

50. Wrasidlo, W. *et al.* A de novo compound targeting α-synuclein improves deficits in models of Parkinson's disease. *Brain* **139**, 3217–3236 (2016).

51. Perni, M. *et al.* A natural product inhibits the initiation of α-synuclein aggregation and suppresses its toxicity. *Proc. Natl. Acad. Sci.* **114**, E1009–E1017 (2017).

52. Lee, E.-N. *et al.* Phthalocyanine tetrasulfonates affect the amyloid formation and cytotoxicity of alpha-synuclein. *Biochemistry* **43**, 3704–15 (2004).
53. Fonseca-Ornelas, L. *et al.* Small molecule-mediated stabilization of vesicle-associated helical α-synuclein inhibits pathogenic misfolding and aggregation. *Nat. Commun.* **5**, 5857 (2014).
54. Danzer, K. M. *et al.* Heat-shock protein 70 modulates toxic extracellular α-synuclein oligomers and rescues trans-synaptic toxicity. *FASEB J.* **25**, 326–336 (2011).
55. Moree, B. *et al.* Small molecules detected by second-harmonic generation modulate the conformation of monomeric α-synuclein and reduce its aggregation in cells. *J. Biol. Chem.* **290**, 27582–27593 (2015).
56. Collier, T. J. *et al.* Nortriptyline inhibits aggregation and neurotoxicity of alpha-synuclein by enhancing reconfiguration of the monomeric form. *Neurobiol. Dis.* **106**, 191–204 (2017).
57. Prabhudesai, S. *et al.* A Novel 'Molecular Tweezer' Inhibitor of α-Synuclein Neurotoxicity *in Vitro* and *in Vivo*. *Neurotherapeutics* **9**, 464–476 (2012).
58. Acharya, S. *et al.* Molecular basis for preventing α-synuclein aggregation by a molecular tweezer. *J. Biol. Chem.* **289**, 10727–10737 (2014).
59. Burré, J., Sharma, M. & Südhof, T. C. α-Synuclein assembles into higher-order multimers upon membrane binding to promote SNARE complex formation. *Proc. Natl. Acad. Sci.* **111**, E4274–E4283 (2014).
60. Dettmer, U. *et al.* Loss of native α-synuclein multimerization by strategically mutating its amphipathic helix causes abnormal vesicle interactions in neuronal cells. *Hum. Mol. Genet.* **26**, 3466–3481 (2017).
61. Lou, X., Kim, J., Hawk, B. J. & Shin, Y.-K. α-Synuclein may cross-bridge v-SNARE and acidic phospholipids to facilitate SNARE-dependent vesicle docking. *Biochem. J.* **474**, 2039–2049 (2017).
62. Caruana, M. *et al.* Inhibition and disaggregation of α-synuclein oligomers by natural polyphenolic compounds. *FEBS Lett.* **585**, 1113–1120 (2011).
63. Asiamah, I., Hodgson, H. L., Maloney, K., Allen, K. J. H. & Krol, E. S. Ring substitution influences oxidative cyclisation and reactive metabolite formation of nordihydroguaiaretic acid analogues. *Bioorganic Med. Chem.* **23**, 7007–7014 (2015).
64. Mazzulli, J. R., Burbulla, L. F., Krainc, D. & Ischiropoulos, H. Detection of Free and Protein-Bound ortho-Quinones by Near-Infrared Fluorescence. *Anal. Chem.* **88**, 2399–405 (2016).
65. Natalello, A., Santambrogio, C. & Grandori, R. Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? *J. Am. Soc. Mass Spectrom.* **28**, 21–28 (2017).
66. Ferrie, J. J. *et al.* Using a FRET Library with Multiple Probe Pairs To Drive Monte Carlo Simulations of α-Synuclein. *Biophys. J.* **114**, 53–64 (2018).
67. George, J. M., Jin, H., Woods, W. S. & Clayton, D. F. Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. *Neuron* **15**, 361–372 (1995).
68. Davidson, W. S., Jonas, A., Clayton, D. F. & George, J. M. Stabilization of α-Synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443–9449 (1998).
69. Ulmer, T. S. & Bax, A. Comparison of structure and dynamics of micelle-bound human α-synuclein and Parkinson disease variants. *J. Biol. Chem.* **280**, 43179–43187 (2005).
70. Rhoades, E., Ramlall, T. F., Webb, W. W. & Eliezer, D. Quantification of α-synuclein binding to lipid vesicles using fluorescence correlation spectroscopy. *Biophys. J.* **90**, 4692–4700 (2006).
71. Martinez, B. A., Caldwell, K. A. & Caldwell, G. A. *C. elegans* as a model system to accelerate discovery for Parkinson disease. *Curr. Opin. Genet. Dev.* **44**, 102–109 (2017).
72. Hamamichi, S. *et al.* Hypothesis-based RNAi screening identifies neuroprotective genes in a Parkinson's disease model. *Proc. Natl. Acad. Sci. USA* **105**, 728–733 (2008).
73. Zharikov, A. D. *et al.* shRNA targeting α-synuclein prevents neurodegeneration in a Parkinson's disease model. *J. Clin. Invest.* **125**, 2721–2735 (2015).
74. Lee, H.-J. Clearance of α-Synuclein Oligomeric Intermediates via the Lysosomal Degradation Pathway. *J. Neurosci.* **24**, 1888–1896 (2004).
75. Dehay, B. *et al.* Pathogenic Lysosomal Depletion in Parkinson's Disease. *J. Neurosci.* **30**, 12535–12544 (2010).
76. Masliah, E. *et al.* Effects of alpha-synuclein immunization in a mouse model of Parkinson's disease. *Neuron* **46**, 857–68 (2005).
77. Masliah, E. *et al.* Passive immunization reduces behavioral and neuropathological deficits in an alpha-synuclein transgenic model of lewy body disease. *PLoS One* **6** (2011).
78. Levin, J. *et al.* The PROMESA-protocol: progression rate of multiple system atrophy under EGCG supplementation as anti-aggregation-approach. *J. Neural Transm.* **123**, 439–45 (2016).
79. Reznichenko, L., Kalfon, L., Amit, T., Youdim, M. B. H. & Mandel, S. A. Low dosage of rasagiline and epigallocatechin gallate synergistically restored the nigrostriatal axis in MPTP-induced parkinsonism. *Neurodegener. Dis.* **7**, 219–231 (2010).
80. Chen, M. *et al.* Tea polyphenols alleviate motor impairments, dopaminergic neuronal injury, and cerebral α-synuclein aggregation in MPTP-intoxicated parkinsonian monkeys. *Neuroscience* **286**, 383–392 (2015).
81. Thomas, B. *et al.* Resistance to MPTP-Neurotoxicity in α-synuclein knockout mice is complemented by human α-synuclein and associated with increased β-synuclein and Akt activation. *PLoS One* **6** (2011).
82. Bieschke, J. *et al.* EGCG remodels mature α-synuclein and amyloid-β fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci.* **107**, 7710–7715 (2010).
83. Yang, J. E. *et al.* EGCG-mediated Protection of the Membrane Disruption and Cytotoxicity Caused by the 'Active Oligomer' of α-Synuclein. *Sci. Rep.* **7**, 1–10 (2017).
84. Pogačnik, L. *et al.* Potential for brain accessibility and analysis of stability of selected flavonoids in relation to neuroprotection *in vitro*. *Brain Res.* **1651**, 17–26 (2016).
85. Lambert, J. D. *et al.* Hepatotoxicity of high oral dose (−)-epigallocatechin-3-gallate in mice. *Food Chem. Toxicol.* **48**, 409–416 (2010).
86. Zhu, M. & Fink, A. L. Lipid binding inhibits α-synuclein fibril formation. *J. Biol. Chem.* **278**, 16873–16877 (2003).
87. Burré, J. *et al.* α-Synuclein promotes SNARE-complex assembly *in vivo* and *in vitro*. *Science* **329**, 1663–1667 (2010).
88. Kim, S. *et al.* GBA1 deficiency negatively affects physiological α-synuclein tetramers and related multimers. *Proc. Natl. Acad. Sci.* **115**, 201700465 (2018).
89. Dettmer, U. *et al.* Parkinson-causing α-synuclein missense mutations shift native tetramers to monomers as a mechanism for disease initiation. *Nat. Commun.* **6**, 7314 (2015).
90. Rasia, R. M. *et al.* Structural characterization of copper(II) binding to α-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proc. Natl. Acad. Sci.* **102**, 4294–4299 (2005).
91. Binolfi, A. *et al.* Interaction of α-synuclein with divalent metal ions reveals key differences: A link between structure, binding specificity and fibrillation enhancement. *J. Am. Chem. Soc.* **128**, 9893–9901 (2006).
92. Appel-Cresswell, S. *et al.* Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. *Mov. Disord.* **28**, 811–813 (2013).
93. Proukakis, C. *et al.* A novel α-synuclein missense mutation in Parkinson disease. *Neurology* **80**, 1062–1064 (2013).
94. Ghosh, D. *et al.* The parkinson's disease-associated H50Q mutation accelerates α-synuclein aggregation *in vitro*. *Biochemistry* **52**, 6925–6927 (2013).

95. Kowalik-Jankowska, T., Rajewska, A., Jankowska, E. & Grzonka, Z. Copper(II) binding by fragments of alpha-synuclein containing M1-D2- and -H50-residues; a combined potentiometric and spectroscopic study. *Dalton Trans.* 5068–76, https://doi.org/10.1039/b610619f (2006).

96. Back, P. *et al.* Exploring real-time *in vivo* redox biology of developing and aging Caenorhabditis elegans. *Free Radic. Biol. Med.* **52**, 850–859 (2012).

97. Hammoudeh, D. I., Follis, A. V., Prochownik, E. V. & Metallo, S. J. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.* **131**, 7390–401 (2009).

98. Metallo, S. J. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* **14**, 481–8 (2010).

99. Dunker, A. K. & Uversky, V. N. Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* **10**, 782–8 (2010).

100. Yin, X., Giap, C., Lazo, J. S. & Prochownik, E. V. Low molecular weight inhibitors of Myc-Max interaction and function. *Oncogene* **22**, 6151–9 (2003).

101. Wang, H. *et al.* Improved low molecular weight Myc-Max inhibitors. *Mol. Cancer Ther.* **6**, 2399–408 (2007).

102. Theillet, F. X. *et al.* Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).

103. Winner, B. *et al. In vivo* demonstration that α-synuclein oligomers are toxic. *Proc. Natl. Acad. Sci.* **108**, 4194–4199 (2011).

104. Favier, A. & Brutscher, B. Recovering lost magnetization: Polarization enhancement in biomolecular NMR. *J. Biomol. NMR* **49**, 9–15 (2011).

105. Brutscher, B. *et al.* NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv. Exp. Med. Biol.* **870**, 49–122 (2015).

106. Huang, C., Ren, G., Zhou, H. & Wang, C. C. A new method for purification of recombinant human α-synuclein in Escherichia coli. *Protein Expr. Purif.* **42**, 173–177 (2005).

107. Berkowitz, L. A., Knight, A. L., Caldwell, G. A. & Caldwell, K. A. Generation of Stable Transgenic *C. elegans* Using Microinjection. *J. Vis. Exp.* 3–5, https://doi.org/10.3791/833 (2008).

Acknowledgements

Author Contributions

M.D. designed, conducted, and analyzed the α-synuclein aggregation experiments. M.D. analyzed all data and wrote the manuscript. J.B.N., H.K., K.C. and G.C. designed, conducted, analyzed, and wrote description of the *C. elegans* experiments and assisted in editing the manuscript. V.S., M.S., M.G.M., R.P. and I.F. designed, conducted, analyzed, and wrote description of the N.M.R. experiments and assisted in editing the manuscript. B.P. conducted, analyzed, and wrote description of the FCS. experiments and assisted in editing the manuscript. J.J.F. conducted, analyzed, and wrote description of the FRET experiments. C.H. generated some of the labeled α-synuclein used in the FRET experiments. E.R. and E.J.P. assisted in design and analysis of the FCS and FRET experiments and assisted in editing the manuscript. R.M. and F.S. conducted, analyzed, and wrote description of the IM-MS experiments and assisted in editing the manuscript. A.N. and R.G. conducted, analyzed, and wrote description of the ESI-MS experiments and assisted in editing the manuscript. N.G. and V.U. assisted in writing and editing the manuscript. E.K. generated the cyclized NDGA analogs and assisted in experimental design, writing, and editing of the manuscript. H.I. planned, organized, and participated in the analysis and discussion of all the data and editing of the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39480-z.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
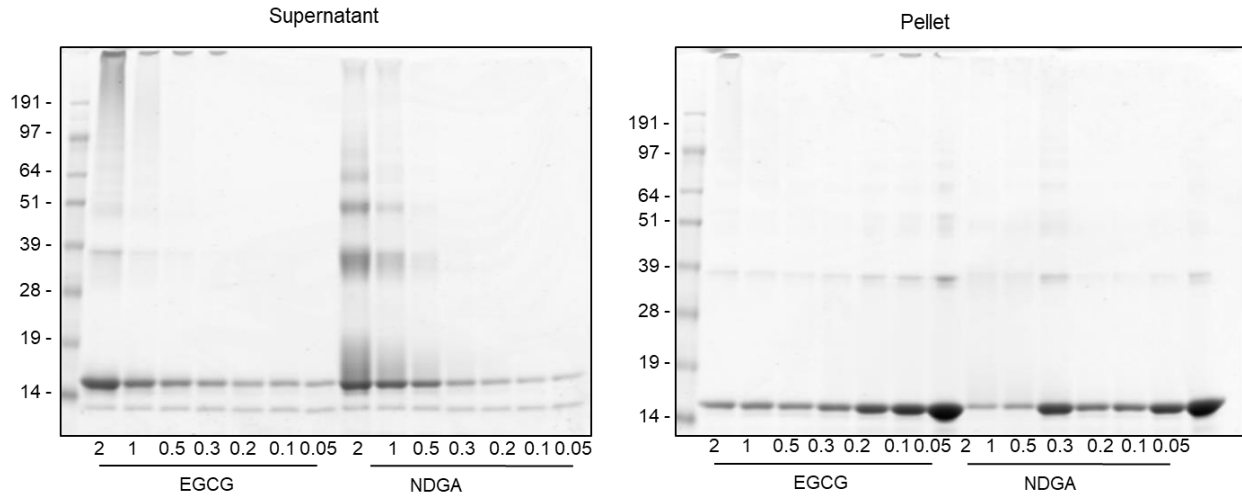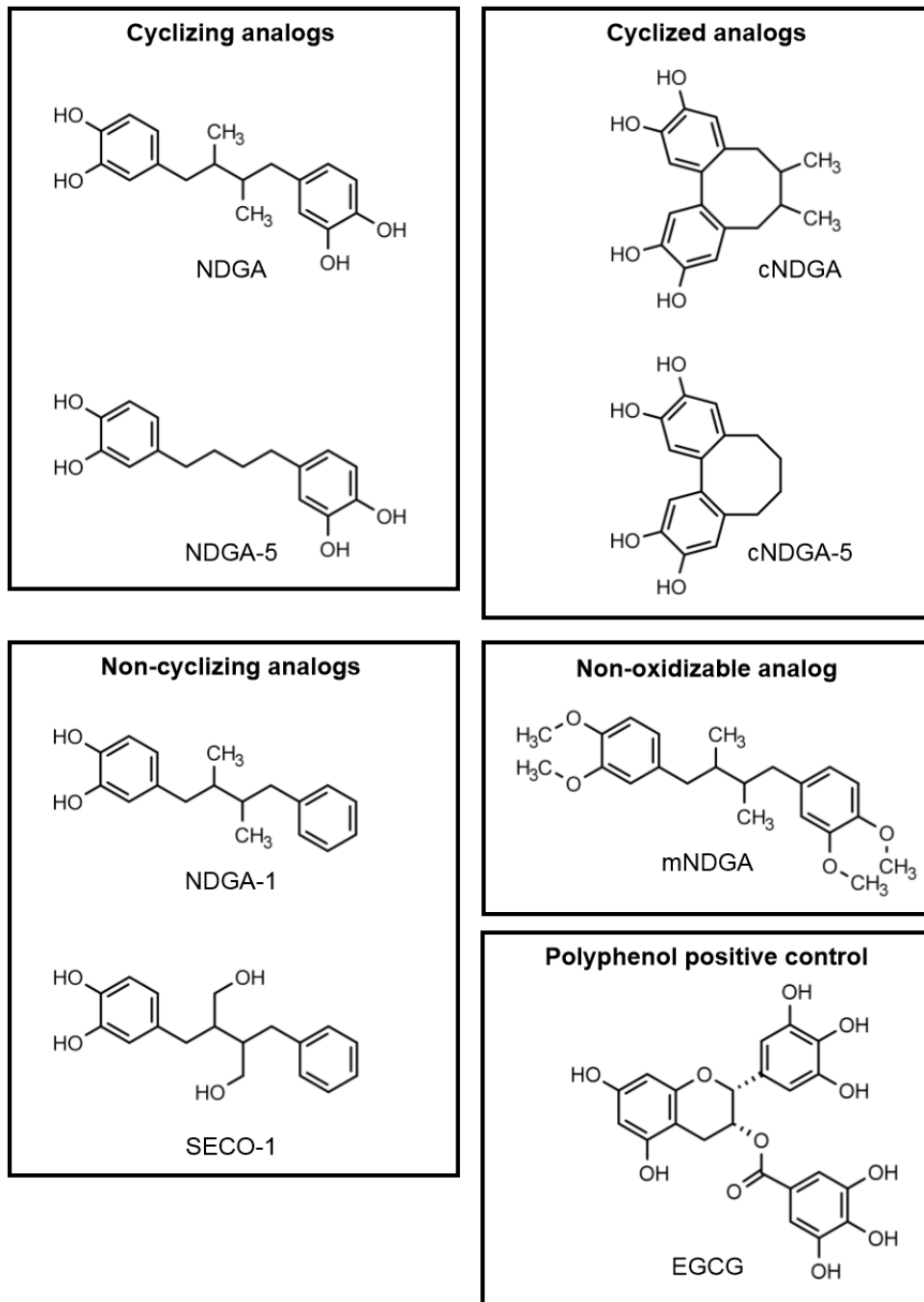
Supplementary Figures

Title: Cyclized NDGA modifies dynamic α-synuclein monomers preventing aggregation and toxicity

Authors: Malcolm J. Daniels, J. Brucker Nourse Jr., Hanna Kim, Valerio Sainati, Marco Schiavina, Maria Grazia Murrali, Buyan Pan, John J. Ferrie, Conor M. Haney, Rani Moons, Neal S. Gould, Antonino Natalello, Rita Grandori, Frank Sobott, E. James Petersson, Elizabeth Rhoades, Roberta Pierattelli, Isabella Felli, Vladimir N. Uversky, Kim A. Caldwell, Guy A. Caldwell, Edward S. Krol and Harry Ischiropoulos

Supplementary Figure S1. Dose effect of EGCG or NDGA on α-synuclein aggregation. Representative images of recombinant human wildtype α-synuclein (138 µM) aggregated for 7 days in the presence of EGCG or NDGA at the indicated molar ratios. After aggregation, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble (supernatant) and insoluble (pellet) fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry. (n = 3-5).

Supplementary Figure S2. Structures of EGCG, NDGA, and NDGA analogs employed in this study.
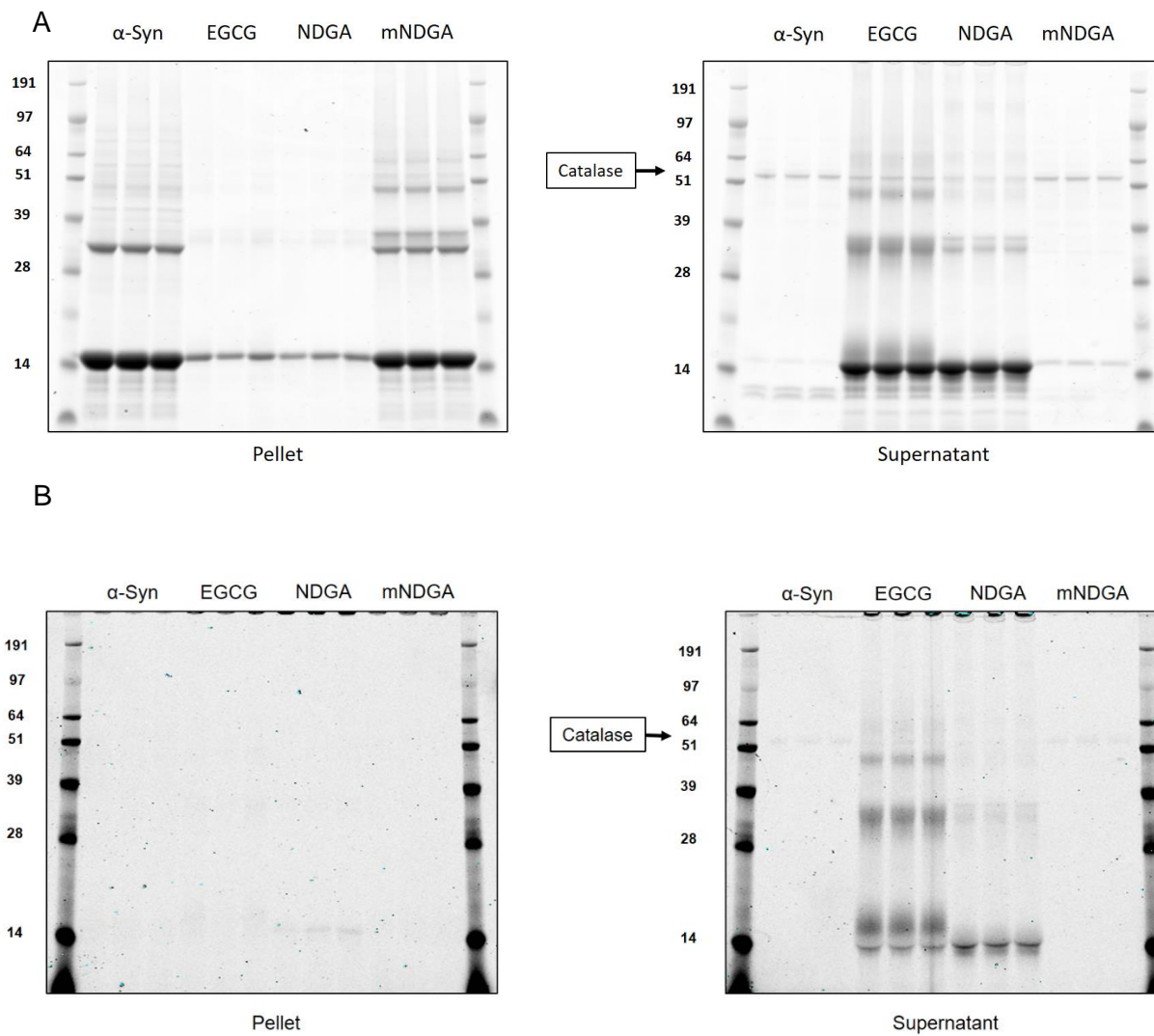
Supplementary Figure S3. Colloidal and nIRF images of α-synuclein aggregation in the presence of 5% catalase. **(a)** Colloidal stained gel **(b)** nIRF image of the same gel before colloidal staining, n=3.
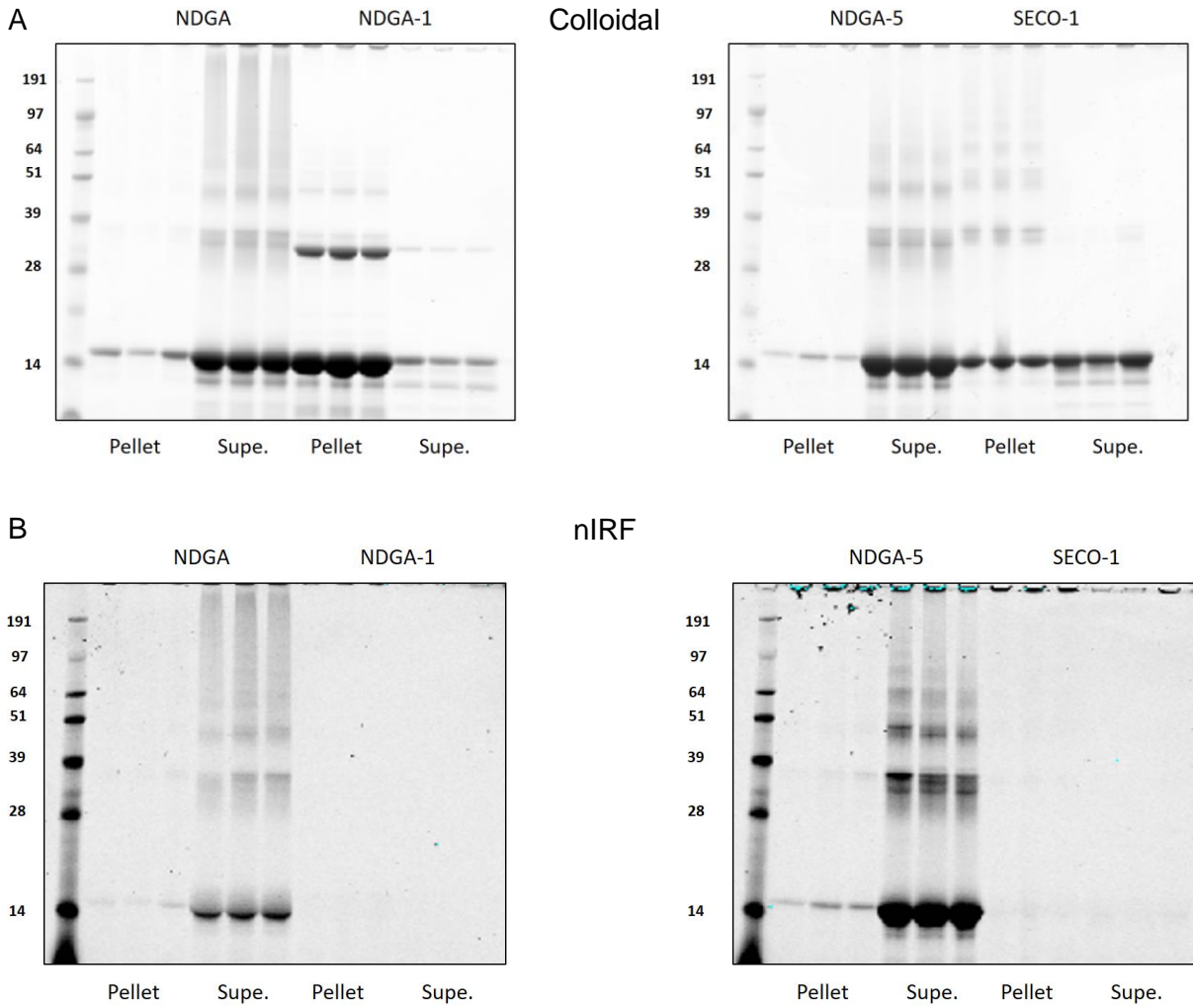
Supplementary Figure S4. Colloidal and nIRF images of α-synuclein aggregation in the presence of NDGA analogs. **(a)** Colloidal images of gels after separation by centrifugation. **(b)** Near-infrared images of the same gels before colloidal staining, n=3.
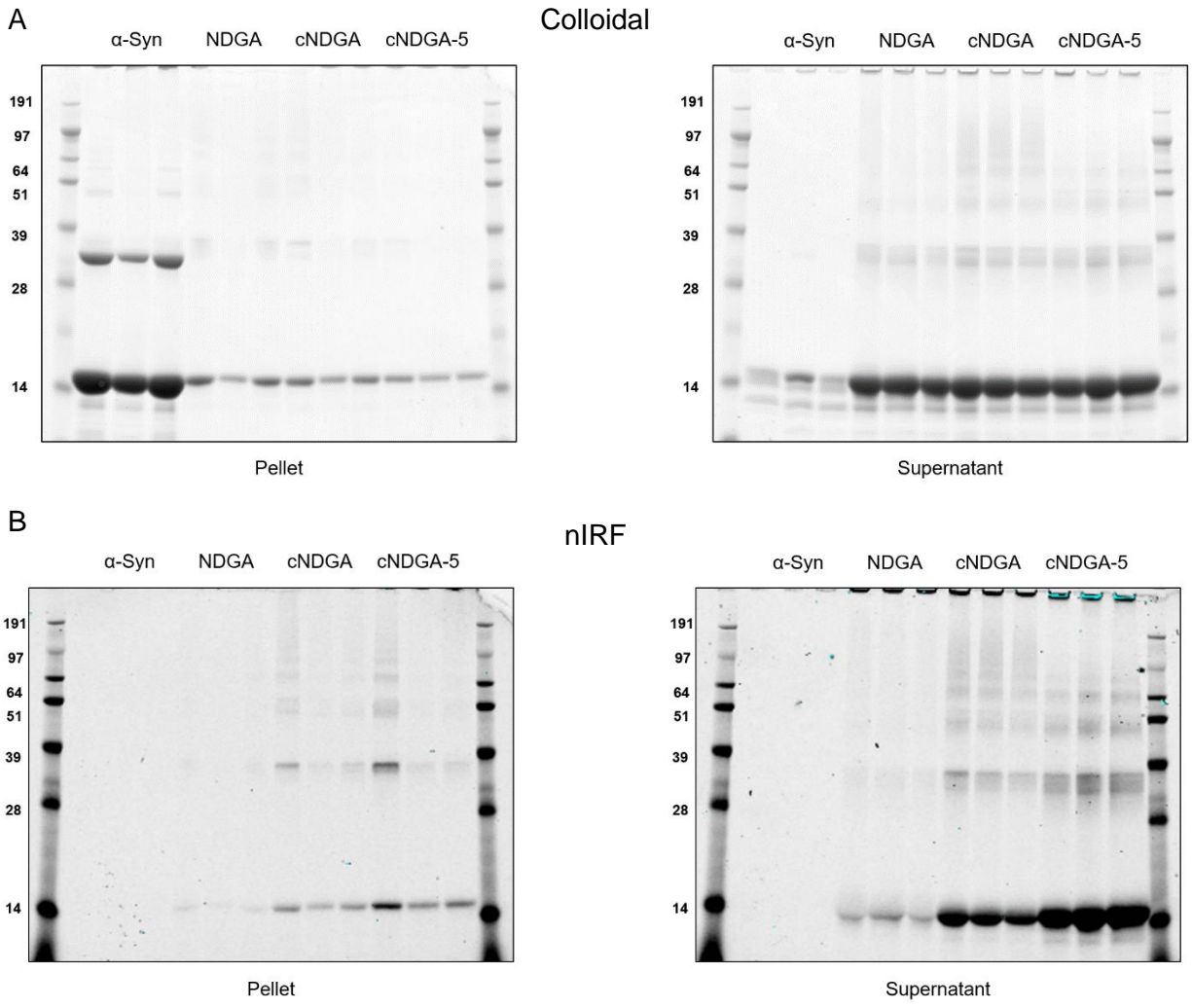
Supplementary Figure S5. Colloidal and nIRF images of α-synuclein aggregation in the presence of cyclized NDGA analogs. **(a)** Colloidal images of gels after separation by centrifugation. **(b)** Near-infrared images of the same gels before colloidal staining, n=3.
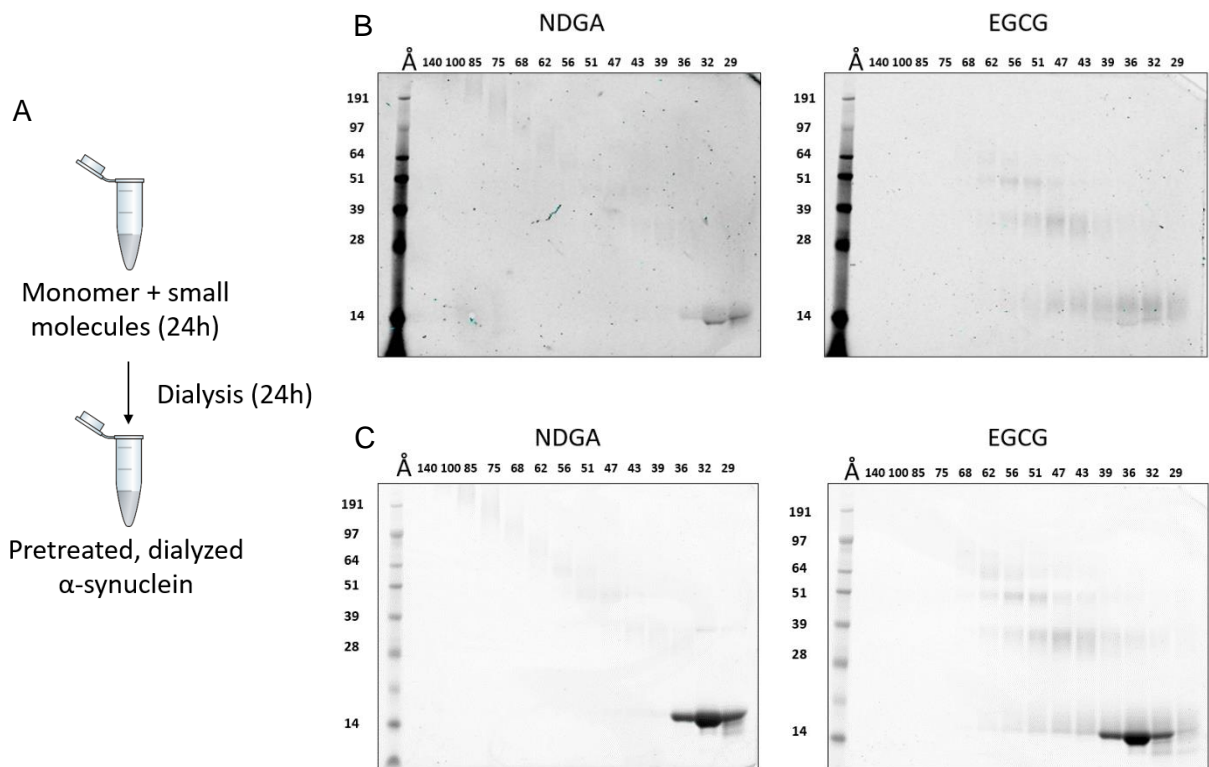
Supplementary Figure S6. NDGA pretreatment produces nIRF positive α-synuclein monomers.

(a) Workflow for CD analysis and aggregation. **(b)** Near-infrared fluorescence images of NDGA and EGCG pretreated, dialyzed α-synuclein separated by native-state size-exclusion chromatography. **(c)** Colloidal staining of the same gels shown in panel b.
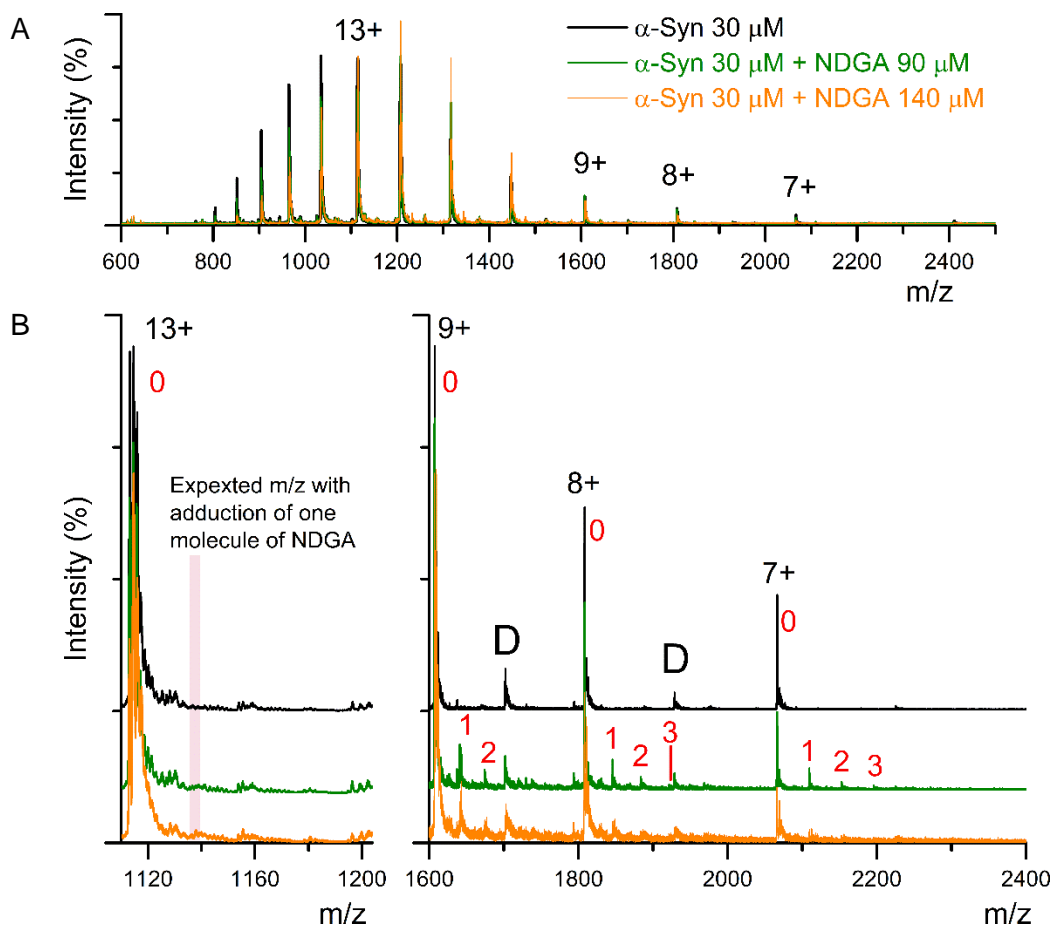
Supplementary Figure S7. Electrospray mass spectrometry characterization of α-synuclein-NDGA interaction. (a) Nano-ESI-MS spectra in positive-ion mode of 30 μM α-synuclein in the absence and presence of NDGA at 90 μM and 140 μM. (b) Magnification of the spectra in the Upper panel. The number of NDGA molecules bound to α-synuclein monomers is indicated by red numbers. Dimer-specific peaks of α-synuclein are labeled as "D". Charge states are indicated by black numbers. Nano-ESI-MS spectra were collected as described in Konijnenberg et al.[49]
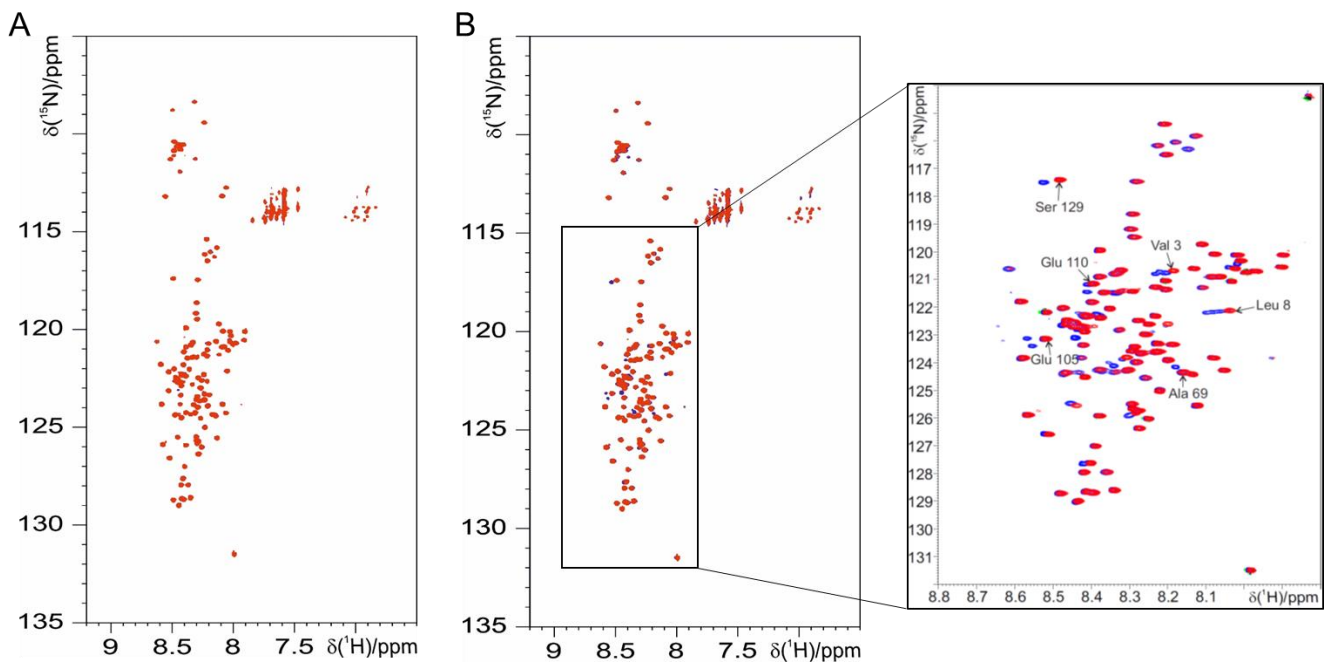
Supplementary Figure S8. 2D NMR spectra comparing α-synuclein treated with 1:1 NDGA for 24 hours (red), with **(a)** α-synuclein treated with 3:1 NDGA for 24 hours (blue), and **(b)** α-synuclein treated with 1:1 NDGA for 10 days (blue).
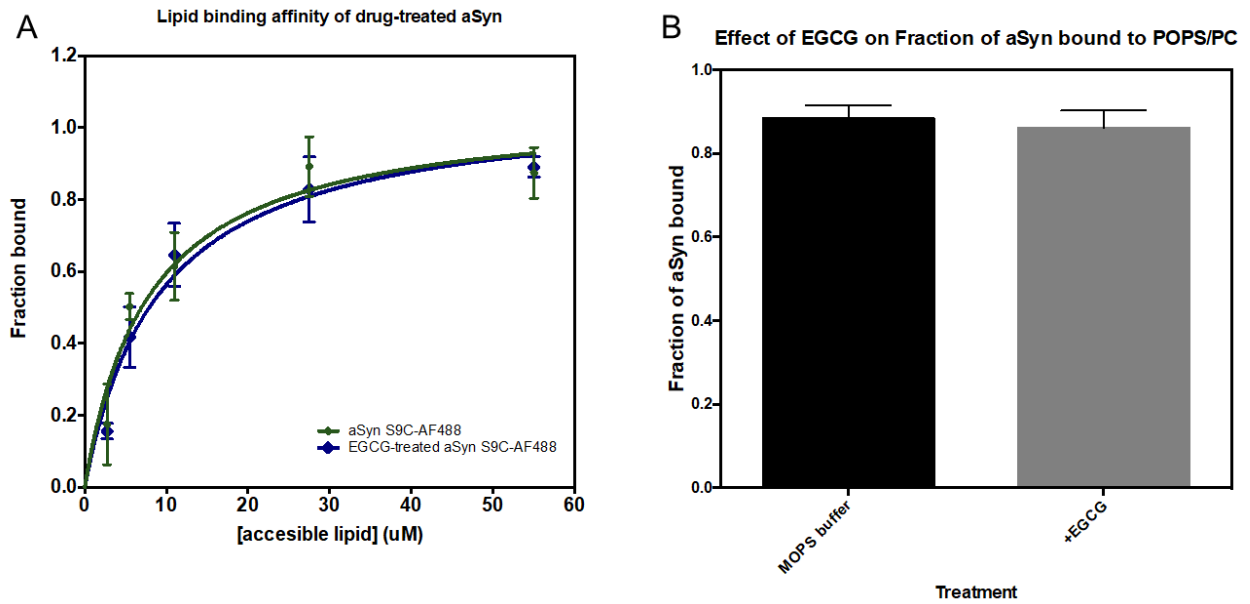
Supplementary Figure S9. EGCG does not disrupt lipid binding. (A) α-Synuclein was incubated 1:1 with EGCG or solvent alone for 24 hours before fluorescence correlation spectroscopy in the presence of POPS:POPC vesicles at the indicated concentrations. (n = 3). (B) Addition of NDGA did not displace fluorescently labeled α-synuclein from POPS:POPC vesicles. (n = 3).
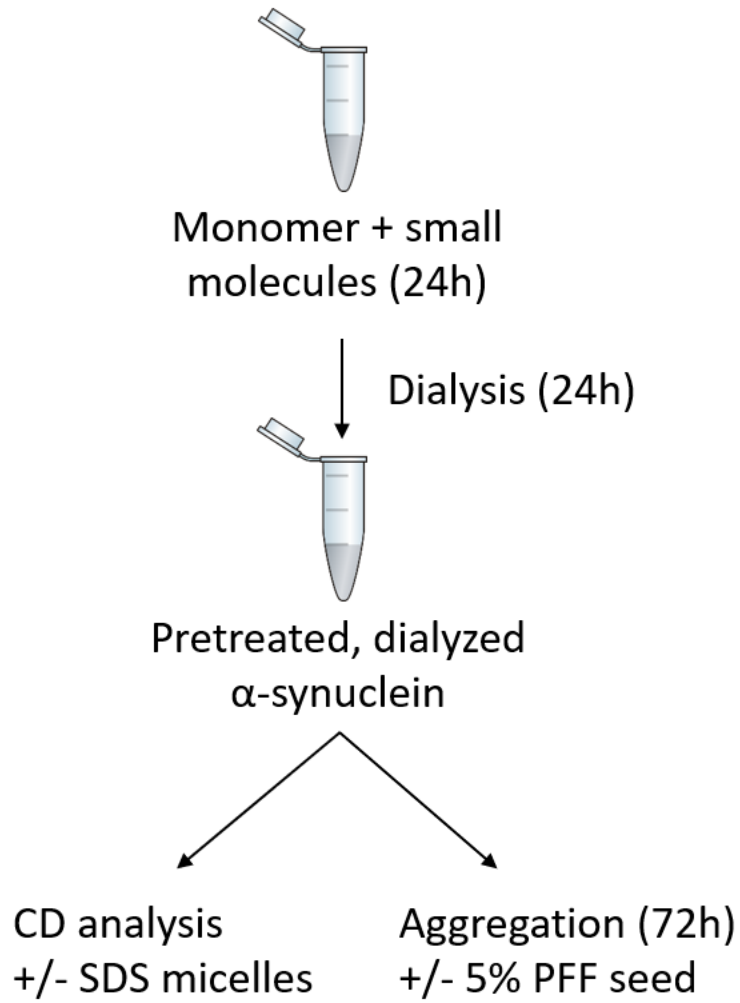
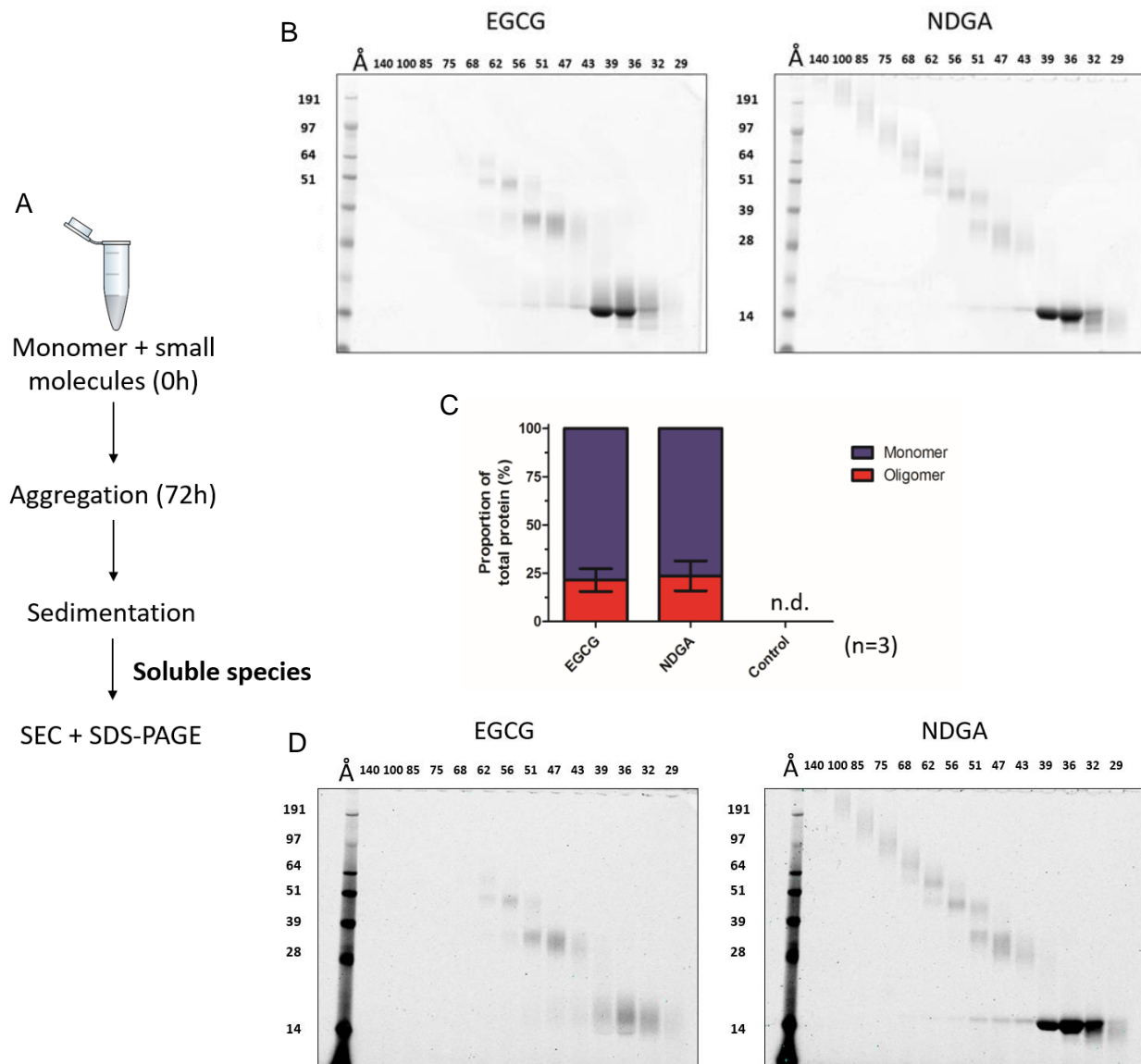Supplementary Figure S10. Workflow for analysis of pretreated, dialyzed α-synuclein
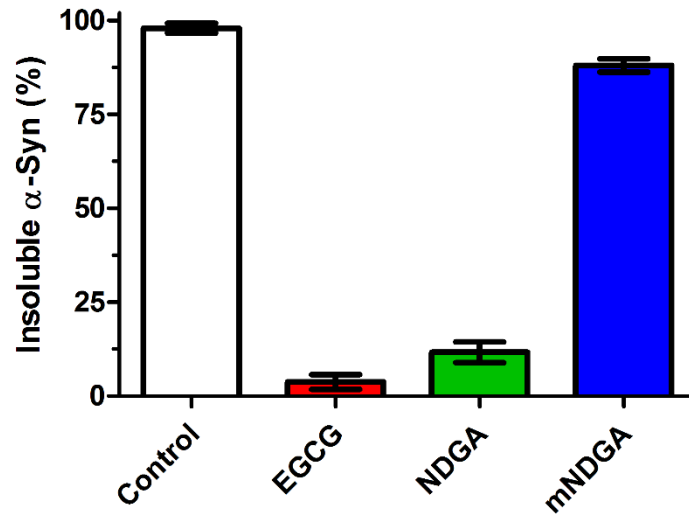
Supplementary Figure S11. NDGA pretreatment inhibits α-synuclein aggregation despite 14 days under aggregation conditions. α-Synuclein was incubated 1:1 with small molecules for 24 hours then dialyzed against PBS for 24 hours. After aggregation for 14 days, PBS-insoluble α-synuclein was separated by centrifugation (21k g for 10 min). Soluble and insoluble fractions were boiled in SDS, run by SDS-PAGE, and colloidal stained. α-Synuclein in each fraction was quantified by in-gel densitometry. (n = 3)
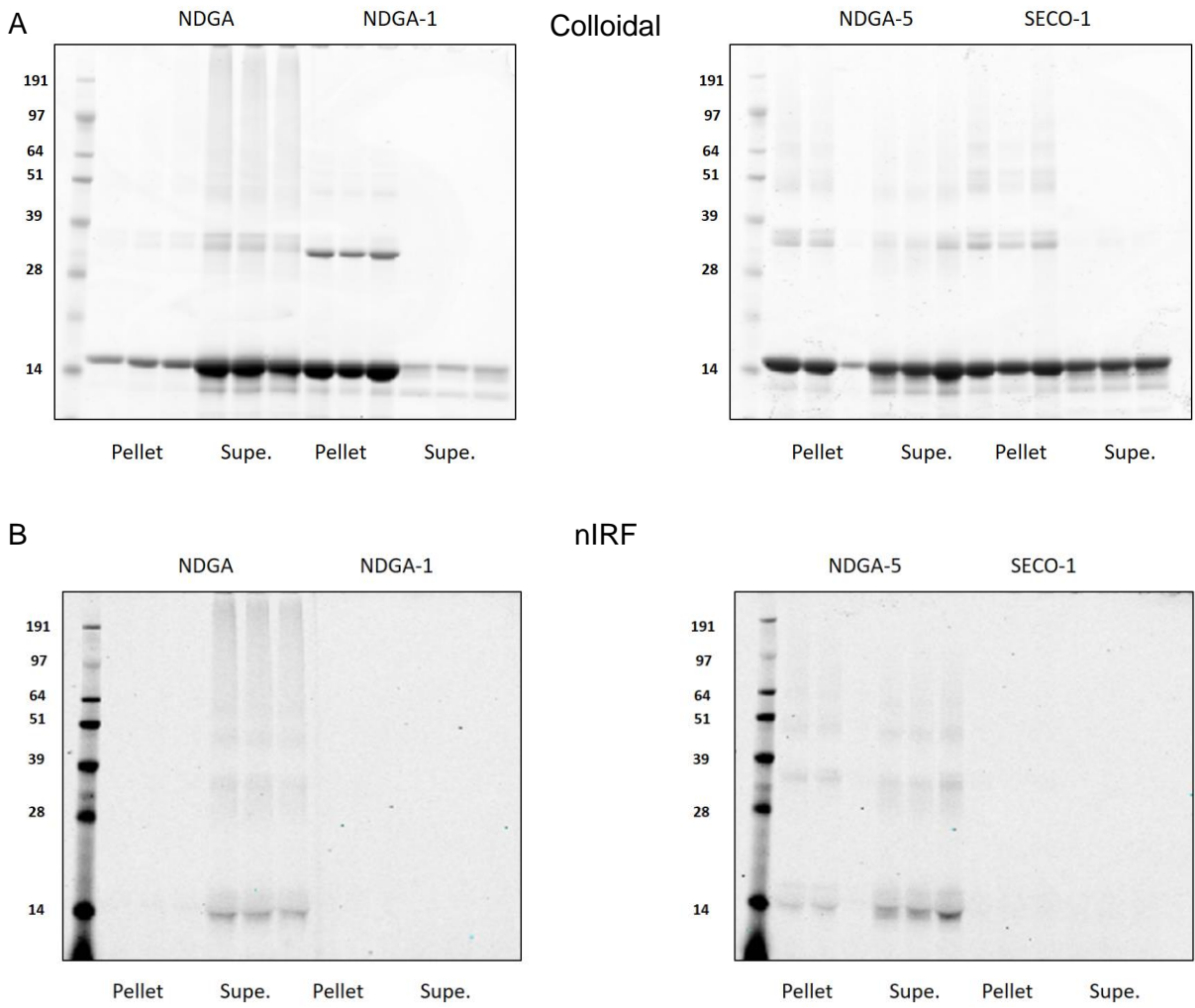
Supplementary Figure S12. Colloidal and nIRF images of NDGA analog pretreated, dialyzed α-synuclein aggregation **(a)** Colloidal images of gels after separation by centrifugation. **(b)** Near-infrared images of the same gels before colloidal staining, n=3.
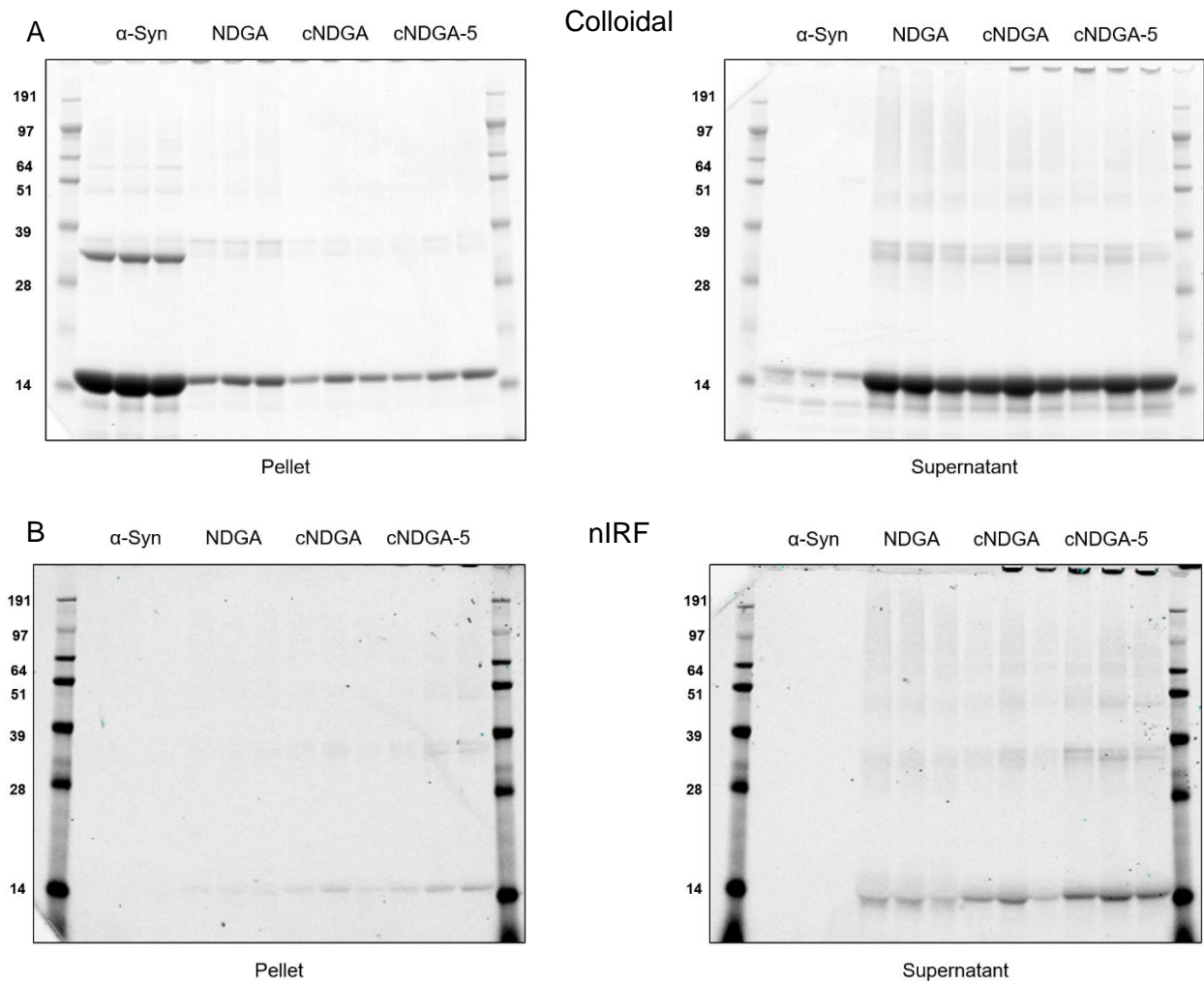
Supplementary Figure S13. Colloidal and nIRF images of cyclized NDGA analog pretreated, dialyzed α-synuclein aggregation **(a)** Colloidal images of gels after separation by centrifugation. **(b)** Near-infrared images of the same gels before colloidal staining, n=3.

Supplementary Figure S14. nIRF positive monomers are the predominant product of α-synuclein aggregation in the presence of EGCG and NDGA. (a) Workflow (b) Colloidal (c) Densitometry (d) nIRF (n=3).
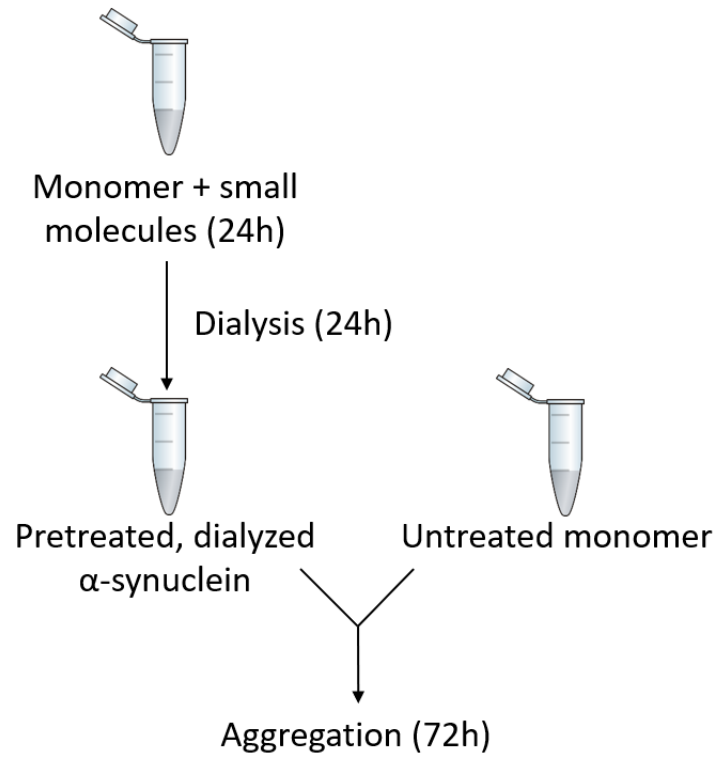
Supplementary Figure S15. Workflow for aggregation of combined mixtures of pretreated, dialyzed α-synuclein and untreated α-synuclein.
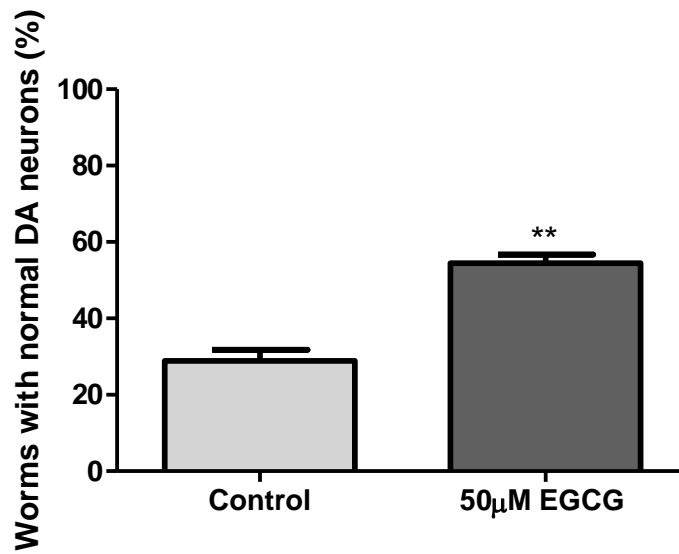
Supplementary Figure S16. EGCG reduces α-synuclein-driven neurodegeneration. *C. elegans* were treated with vehicle (dH$_2$O) or 50μM EGCG in NGM agar on days 0-3, 5, and 7. Dopamine neurons were scored on day 8. Worms were scored as normal if no degenerative phenotypes (broken dendritic process, cell body loss, dendritic blebbing, or a missing neuron) were observed. Data represented as mean ± SEM.; n=3 replicates with 30 worms per replicate; t-test; **p= 0.0023

3.2 Structural characterization of a large IDPs

The Nipah virus (NiV), together with its close relative Hendra virus (HeV), has been identified within the *Paramyxoviridae* family as a zoonotic agent responsible for severe encephalitis in humans. Because of their high pathogenicity and the lack of therapeutics and vaccines, Henipaviruses are classified among bio-security level 4 (BSL-4) pathogens and are considered as potential bio-terrorism agents. One of the most important protein for transcription and replication of these viruses is the P protein. In addition, the P gene also encodes for the W and the V proteins, and these three proteins are all identical in their N – terminal region (aa. 1-406). The Henipavirus P, V and W proteins play a key role in the evasion of the Interferon (IFN) mediated response through an antagonist activity of IFN signaling. Thus, the structural characterization of these proteins, and of the P N-terminal (PNT) in particular, can be a starting point towards the design of inhibitors of this deadly virus.

The high level of disorder and flexibility rendered the assignment of this 406 amino acid long protein quite challenging. The use of ^{13}C detection was crucial to overcome the little dispersion of the NMR signals and the large extent of overlaps in ^{1}H detected experiments. The use of multidimensional heteronuclei experiments, allowed to assign the resonances of HN, N, C', C$^\alpha$, and C$^\beta$ nuclei of 352 residues, including 20 out of the 26 prolines, thus leading to a percentage of 85% of assigned residues for the entire protein.

Of course, the assignment of a large disordered protein like this is not easy and requires to use a large number of experiments and a sufficient amount of ^{13}C-^{15}N labeled protein to exploit ^{13}C multidimensional experiments. Nevertheless, this is still the most accurate and straightforward way to obtain the structural characterization of proteins like this.

The APSY spectroscopy, previously described in Chapter 2 was used to obtain a first assignment of the backbone, using a 5D (HCA)CONCACON experiment complemented by a 3D HNCO. The assignment was extended with a combination of ^{1}H-detected and ^{13}C-detected experiments, including also C$^\beta$ nuclei. Exploiting the heteronuclear chemical shift of all the assigned signals it was possible to calculate the neighbor-corrected structure propensities (ncSSP)[100,101] to evaluate

the presence of residual secondary structure. The backbone dynamics was investigated measuring the ^{15}N transverse and longitudinal relaxation rates (^{15}N R_2 and ^{15}N R_1) and ^{1}H-^{15}N NOE. These three observables provide information about the protein mobility, elucidating the different level of flexibility within the whole protein.

The data obtained from NMR experiments will be combined with SAXS experiments, to obtain a conformational ensemble able to describe the conformational behavior of NiV PNT in solution.

Above the importance of the functional aspect of the structural characterization of this viral protein, this study provides another proof of the importance of NMR spectroscopy to investigate disordered proteins. Combination of NMR with other techniques is of course also very relevant to obtain a complete description of a macromolecule. In particular, SAXS is not limited by size, yields overall shapes and it is sensitive to domain/subunit positions, whereas NMR provides high resolution models and is sensitive to domain/subunit orientations. The complementarity of orientational and short-range distance data from NMR and the long-range distance and shape information from SAXS, is particularly important for highly flexible system for which X-ray crystallography cannot provide any information.

3.2.1 Article: Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS

Ensemble description of the intrinsically disordered N-terminal domain of the Nipah virus P/V protein from combined NMR and SAXS

Edoardo Salladini[1], Maria Grazia Murrali[2], Christophe Bignon[1], Isabella C. Felli[2], Roberta Pierattelli[2] and Sonia Longhi[1]

[1]Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Marseille, France

[2]Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

In preparation

Abstract

Using SAXS and NMR spectroscopy, we herein provide a quasi-atomic resolution description of the spectacularly large (aa 1-406) intrinsically disordered N-terminal domain (PNT) of the Nipah virus (NiV) phosphoprotein (P) and V protein, two key players in viral genome replication and in evasion of the host innate immune response, respectively. The use of multidimensional NMR spectroscopy allowed us to assign as much as 85 % of the peaks of this long intrinsically disordered domain. The N-terminal region of the protein is the one with the poorest assignment, reflecting the occurrence of a transiently populated α-helix that folds upon binding to the monomeric form of the nucleoprotein. In spite of the mostly disordered nature of PNT, chemical shifts and relaxation measurements show the presence of a few transiently structured elements that correspond to predicted molecular recognition elements (MoREs). The combination of SAXS and NMR restraints, allowed obtaining an ensemble description of the conformational behavior of the protein in solution. The present results beyond providing an overall description of the conformational behavior of this large intrinsically disordered region also constitute an asset for obtaining atomistic information in future interaction studies with viral and/or cellular partners. The present study can thus be regarded as the starting point towards the design of inhibitors that, by targeting crucial protein-protein interactions involving PNT, might be instrumental to combat this deadly virus.

Introduction

The Nipah virus (NiV), together with its close relative Hendra virus (HeV), has been identified within the *Paramyxoviridae* family as a zoonotic agent responsible for severe encephalites in humans. The NiV and HeV have been grouped within the *Henipavirus* genus [1] that also comprises the recently discovered Cedar virus (CedV) [2]. Because of their high pathogenicity, wide host range, interspecies transmission and lack of therapeutics and vaccines, henipaviruses are classified among bio-security level 4 (BSL-4) pathogens and are considered as potential bio-terrorism agents.

The genome of henipaviruses, made of negative-stranded, non-segmented RNA, is encapsidated by the nucleoprotein (N) within a helical nucleocapsid. The nucleocapsid is the substrate used by the polymerase complex during both transcription and replication. These activities are ensured by the viral RNA-dependent RNA polymerase that is made of a complex between the large (L) protein and the phosphoprotein (P). By interacting with both L and the nucleocapsid, the P protein allows the recruitment of L onto the nucleocapsid template. In addition, P also serves as a chaperon for L [3] and N in that it maintains the latter in a monomeric, RNA-free form [4]. Therefore P is a pivotal protein endowed with multiple functions critical for both transcription and replication.

The repertoire of P functions is further expanded by the peculiar coding capacity of the P gene. Indeed, beyond the P protein, the P gene also encodes the V and W proteins that are produced upon addition of either one (protein V) or two (protein W) non-templated guanosines at the editing site of the P messenger. The P, V and W proteins are therefore identical in their N-terminal region (**Figure 1A**). The *Henipavirus* V and W proteins play a key role in the evasion of the interferon (IFN) mediated response *via* an antagonist activity of IFN signaling [5, 6]. V and W bind to STAT1, a key signal transducer in the IFN pathway of antiviral response, through their common N-terminal region (PNT) [7]. Binding of STAT1 by V results in inhibition of STAT1 translocation into the nucleus, while binding by W leads to sequestration of STAT1 in the nucleus [7]. The NiV P protein is endowed with anti-IFN function too, indicating that the common PNT domain is responsible for the antagonist activity.

We previously showed that *Henipavirus* PNT are intrinsically disordered [8]. The P C-terminal region (PCT) consists of alternating disordered and ordered regions (**Figure 1A**). Indeed, PCT spans a disordered region (aa 407-469), referred to as « spacer », a structured region responsible for P oligomerization (PMD, aa 470-578) [9, 10], a disordered linker and a structured region (X domain, XD, aa 660-709) adopting a triple α-helical bundle (**Figure 1A**)

[11]. The C-terminal domain unique to the V protein is predicted to bind zinc and to be folded [12, 13]. We have recently characterized the *Henipavirus* V proteins and showed that PNT remains disordered also in the context of the V protein, while the zinc-finger domain (ZnFD) adopts a predominant β conformation [14]. In that previous study, we also provided the first experimental evidence of the ability of *Henipavirus* V proteins to interact with DDB1 - a cellular protein whose binding to V promotes STAT1 degradation - and showed that the ZnFD plays a crucial role in strengthening this interaction [14].

Although our previous study on the *Henipavirus* V proteins has contributed to illuminate the conformational behavior of these proteins and has provided a conceptual asset to design new antiviral strategies to combat the ability of these viruses to escape the innate immune response, a high-resolution description of these proteins is lacking, as is that of their P proteins. A major hurdle in this respect arises from the presence of their large intrinsically disordered PNT region. High-resolution data are however available for the first 35 residues of NiV PNT. Yabukarski and co-workers solved the X-ray structure of a complex made of a truncated form of NiV N unable to self-assemble (aa 32-383) and of the first 50 residues of NiV P [4] (**Figure 1A**). Only the first 35 residues of P were defined in the electron density. This Molecular Recognition Element (MoRE) forms two α-helices (αP1a, residues 1-19; αP1b; aa 21–28) separated by a kink (**Figure 1A**). In that study, the P region encompassing the first 100 residues (P_{1-100}) was also investigated by NMR spectroscopy on a small construct. Although the HN correlation NMR spectrum is typical of a disordered protein, the secondary-structure propensities (SSPs) calculated from C^{α} and C^{β} secondary chemical shifts indicate the presence of five fluctuating α-helices (referred to as αP1 to αP5) (see **Figure 1B**). Upon addition of the N protein, only residues 50-100 remain detectable indicating that they remain flexible in the complex and that the N°-binding region of P encompasses residues 1 to 50 [4]. This latter region contains two conserved motifs in *Paramyxoviridae* members (i.e. *soyuz1* ad *soyuz2*) [15] (see **Figure 1B**).

As a first step towards an atomic description of the NiV P and V proteins, we herein have investigated NiV PNT by combining small angle X-ray scattering (SAXS) and NMR spectroscopy. The combination of these two approaches allowed us to describe this large intrinsically disordered region as a conformational ensemble and to obtain a better understanding of the dynamic and conformational features of this disordered region.

Results and Discussion

SAXS measurements and calculation of the radius of gyration

The shapes of the SAXS curves obtained at two concentrations of NiV PNT are very similar to each other's, indicating the absence of significant aggregation (**Supplementary Figure S1**). Each curve can be well approximated by a straight line in the Guinier region (qR_g < 1.0). The slope of the curve is proportional to R_g, while the intercept of the straight line gives the forward intensity at 0 angle, I(0), which is proportional to the molecular mass of the scatterer. The R_g values at the two protein concentrations, as calculated by the Guinier plots, are in good agreement with each other's (**Supplementary Figure S1**). We therefore used for subsequent analyses the scattering data obtained at the highest concentration (**Figure 2A**) so as to achieve maximal resolution. The Guiner plot, as obtained from the curve at the highest protein concentration, is shown in **Figure 2B**. The molecular mass determined from the extrapolated scattering intensity at zero angle I(0) is 44.8 kDa, a value close to that expected from the amino acid sequence (45.3 kDa). The distribution of internal distances, as inferred from the scattering curve at the highest protein concentration, yielded a maximal internal dimension D_{max} of 205 Å (**Figure 2C**). This large value and the anisometric nature of the plot indicate that the protein is extended.

A very useful method to describe the structural properties of a molecule is the Kratky plot. In particular, from the shape of this plot one can infer the conformation adopted by the molecule. The Kratky plot of a globular protein has a typical bell shape with a clear maximum. For a completely unfolded protein or for a PMG, no such a maximum can be observed, and the curve displays a plateau. The Kratky plot of NiV PNT (**Figure 2D)** display a plateau for s > 0.7 nm^{-1}. The absence of a maximum clearly indicates that the protein is not globular and does not possess a tightly packed core.

The R_g value experimentally obtained from the curve at the highest concentration (60 Å) was compared with the R_g value expected for a globular, fully unfolded and intrinsically disordered protein of the same size as NiV PNT. The R_g value expected for a globular form is 21.1 Å (see Eq. 2 in Materials and Methods). Therefore, the experimental R_g value is three times higher than expected for a globular form. On the other hand, the expected value for a fully unfolded form is 82.5 Å (see Eq. 3 in Materials and Methods). The R_g value expected for an IDP of the same length as NiV PNT, as calculated using Flory's power law and parameters based on a set of IDPs [16], is 58.8 Å. The experimental R_g value is thus very close, though slightly higher, than the one expected for an IDP of the same size. In agreement with previous

5

findings [17-21], it is conceivable that some specific sequence attributes of NiV PNT, such as the distribution of residues of opposite charge (k value) and/or the proline content, might dictate this slightly more extended conformation. The k value of NiV PNT is relatively low (0.167), and typical of "Janus" sequences, i.e. sequences that can be collapsed or expanded depending on the context.

NMR residue-specific structural and dynamic characterization of NiV PNT

The 2D ^{1}H-^{15}N HSQC spectrum and the 2D CON spectra of NiV PNT (**Figure 3**) confirm that the protein is highly disordered and flexible, as inferred from the small chemical shift dispersion of the ^{1}H NMR signals. It is a very large protein (412 residues including the C-terminal His tag and the initial methionine), which renders its assignment very challenging. In spite of these technical limitations, using multidimensional NMR experiments (**Supplementary Table S1**) we could successfully assign the resonances of H^N, N, C', C$^\alpha$, and C$^\beta$ nuclei for as many as 352 residues, including 20 out of the 26 prolines, thus leading to a percentage of 85% of assigned residues for the entire protein. The assignment, after a final revision, will be deposited in the BMRB. It is worth noting that the assignment is less complete in the N-terminal region of the protein (**Figure 4**), likely reflecting the presence of the relatively long, transiently populated helix αP1 (see **Figure 1B**) [4] that probably leads to an unfavorable conformational exchange regime. When excluding the first 31 residues belonging to helix αP1, the percentage of assigned residues reaches 92%.

The availability of carbon nuclei chemical shifts allows prediction of the presence of residual secondary structure [22]. The neighbor-corrected structure propensities (ncSSPs) [23] were calculated by comparison of the experimentally measured chemical shifts of N, C', C$^\alpha$, and C$^\beta$ nuclei with those of the corresponding random coil values [24]. The propensity to adopt α-helical and β-extended conformations, as obtained using the method of Mulder, are shown in **Figure 4A**.

To investigate the backbone dynamics of NiV PNT, heteronuclear ^{1}H-^{15}N NOE, ^{15}N R_2 and ^{15}N R_1 values were measured. We could successfully determine relaxation values for the majority of the assigned cross-peaks present in the 2D ^{1}H-^{15}N HSQC spectrum, discarding those whose overlap would have made the analysis ambiguous. **Figure 4** shows an overall trend of relaxation values consistent with the predominantly extended conformation of PNT. ^{1}H-^{15}N NOE values are all well below 0.5, indicating high flexibility of the backbone (**Figure 4B**).

Contrary to the ^{15}N R_1 values that are also rather uniform along the polypeptide chain (**Supplementary Figure S2**), meaningfully higher ^{15}N R_2 were observed in some segments (**Figure 4C**).

In addition, CLEANEX experiments were acquired to achieve information on exchange processes of amide protons with the solvent, which are characteristic of solvent exposure (**Figure 4D**). Several experiments were acquired with different mixing times. Intensities of cross peaks, proportional to solvent exposure, are measured and then divided by the analogous values in a reference experiment. Intensity ratios thus report on solvent exchange processes of amide protons on a per-residue basis.

Inspection of the data reported in **Figure 4** allows obtaining a picture of the structural and dynamic properties of NiV PNT. The protein is largely disordered with many regions characterized by mild secondary structural propensies or more compact conformations (grey boxes). The C-terminal part of the protein (aa 320-400) appears quite rich in more structured regions, as evidenced by ncSSP values as well as by higher R_2 rates. The central region of the protein (aa 100-320) is less structured but still shows various short regions characterized by mild structural propensies and higher R_2 rates. Lastly, the N-terminal part could not be fully detected in solution in the full-length construct because many residues in the initial part escape detection due to fast relaxation. This region contains a segment (aa 1-35) whose 3D structure is available in the context of a complex between the first 50 residues of NiV P and a monomeric form of N [4]. The 3D structure of this MoRE reveals two α-helices separated by a kink. These two α-helices could be partially populated in the full-length PNT construct and experience conformational exchange processes that cause broadening of the signals beyond detection. Indeed a few amino acids are detectable and do show nsSP values that indicate a helical propensity (**Figure 4A**). Previous NMR studies carried out on a small construct spanning the first 100 amino acids revealed the presence of five fluctuating helices (αP1 to αP5) [4]. Comparison with the SSPs of the corresponding region within the full-length PNT construct unveils an overall similar profile, with the notable exception of the region spanning residues 37 to 55 (e.g. helix αP2 in the P_{1-100} construct) that does not appear to sample an α-helical conformation in the full-length construct. It is worth noting that the two sets of data were obtained in different experimental conditions (lower T, lower ionic strength and higher pH in the present study than in [4]). Besides, αP2 may partly escape detection due to the incomplete assignment of this region in the context of the full-length construct.

PSIPRED [26] analysis predicts five α-helices within PNT, of which four fall within the region encompassing the first 50 residues and the fifth one encompasses residues 340-348 (**Supplementary Figure S2A**). In addition, PSIPRED also predicts 3 β-strands (aa 233-235, 261-263 and 332-333). Experimental data support the occurrence of the first three helices and of all the β-strands. By contrast, the last predicted α-helix appears to adopt a β conformation (**Figure 4A**). Noteworthy, Milles and co-workers reported a transiently populated α-helix (e.g. helix α4) within the supposedly corresponding region of measles virus (MeV) PNT [27]. It should be pointed out however that aligning intrinsically disordered regions is not an easy procedure, which is further rendered more difficult by the dissimilarity in length between the two PNT domains (230 residues in MeV and 406 residues in NiV). As a result, a certain uncertainty exists as to whether the region spanning helix α4 in MeV PNT is effectively the counterpart of the NiV PNT region encompassing residues 340-348.

In terms of predicted MoREs (see **Supplementary Figure S2B**), a quite good agreement was found with the experimental data with the notable exception of the second MoREs (aa 43-50) that partly overlaps with helix αP2, and which seemingly lacks a transiently populated secondary structure in the context of the full-length PNT construct.

In agreement with previous findings that reported the presence of a STAT1 binding site encompassing residues 50-150 [7], SSP values and relaxation data indicate the presence within the 50-150 region of three transiently structured regions that might fold upon binding to this cellular partner (**Figure 4**). Beyond the STAT1-binding site, analysis of SSPs and relaxation data revealed that the C-terminal moiety of PNT (aa 320-400) likely features at least one additional binding site for one or multiple partners whose identification will require future studies.

Description of NiV PNT as a conformational ensemble

To achieve further insights into the dynamical behavior of NiV PNT, we first assessed the distribution of conformations of the proteins in solution using GAJOE [28]. From an initial pool of 10,000 random conformations, GAJOE selects a sub-ensemble of conformers that collectively reproduces the experimental SAXS data and represents the distribution of structures adopted by the protein in solution. Irrespective of whether the initial ensemble consists of conformers made of Cα-only atoms (as generated by RANCH) or backbone-atoms (as generated by Flexible-Meccano), the average SAXS scattering curve back-calculated from

the selected sub-ensemble reproduces correctly the experimental curve (**Figure 5A**). In both cases, the R_g distribution of the selected sub-ensemble is broad and features a major peak with an R_g value of 64 Å. This peak, while being close to that of the random pool of conformers, is slightly higher, indicating a more extended conformation (**Figure 5B**). Note that successive and independent selections by GAJOE yielded similar results and similar ensembles of conformers showing that the distribution of the selected sub-ensemble was reproducible (data not shown). However, if each individual ensemble matches all experimental data within the error, the different ensembles contain very different structures, because their degrees of freedom largely exceed the number of structural restraints (data not shown).

Next, the data obtained from NMR and SAXS experiments were combined in search of the model conformational ensemble that describes well the conformational behavior of NiV PNT in solution. In a first step, a pool of 10,000 structures was generated based solely on the NMR data. To this end, the ncSSP values were provided as inputs to Flexible-Meccano. Subsequently, from the NMR-based initial pool, a conformational sub-ensemble that best fits the experimental SAXS data was selected using GAJOE with default parameters (**Figure 5C**). When instead of using the default parameters the number of conformers in the final ensemble was varied, the quality of the final fit remained unchanged (data not shown). The R_g distribution of the selected sub-ensemble is less broad than the one obtained without NMR-based structural constraints (**Figure 5D**). Successive and independent selections by GAJOE yielded similar results in terms of R_g distribution, thereby attesting the robustness of the approach (data not shown).

Subsequently, the influence of the occurrence of the N-terminal α-helix in the starting pool of structures on the properties of the final model ensemble was investigated. To this end, using Flexible-Meccano, we generated an initial pool in which an α-helical conformation was imposed to residues 1-30, with experimental ncSSPs (**Figure 5E**). All the resulting sub-ensembles fit equally well the scattering data as judged from the obtained χ^2 values. As the latter were all lower than 1, identifying differences in quality among the various sub-ensembles was not possible. Indeed, if for values greater than 1 the lower χ^2 the more relevant the model, χ^2 values lower than 1 only indicate that the models are equally meaningful according to both standard deviations on intensities and SAXS resolution. In terms of R_g distributions, an increase in the average R_g with the addition of the helix was observed (**Figure 5F**).

SAXS studies unveiled that NiV PNT has an overall extended nature, while being more compact than a fully denatured form. Albeit IDPs/IDRs are notoriously prone to undergo

proteolytic degradation, we managed in generating a NiV PNT sample stable enough to allow acquiring all the NMR spectra required for the assignment. The use of multidimensional NMR spectroscopy allowed us assign as much as 85 % of the peaks of this spectacularly long intrinsically disordered domain. The N-terminal region of the protein is the one with the poorest assignment, reflecting the occurrence of a transiently populated α-helix that folds upon binding to the monomeric form of the N protein. In spite of the overall disordered nature of NiV PNT, the protein samples a few transiently populated regular secondary structure elements. In particular the C-terminal part of the protein (aa 320-400) appears to be enriched in more structured regions that may correspond to MoREs that would fold upon binding to the target.

Conclusion

The combination of SAXS and NMR data led to ensemble models of NiV PNT showing an increased extent of expansion compared to models generated without including secondary structure constraints. The reliability of the inferred structural information, obtained on the isolated PNT domain rather than on the full-length V protein, is supported by previous findings that showed that PNT and the ZnFD do not impact each other's structure and that PNT retains its overall disordered nature also in the context of the V protein [14].

The presence of such large disordered region in the context of the P/V proteins is likely related to their promiscuity. Structural disorder is known to serve as a determinant of protein interactivity [29-31]. Indeed V interacts with various cellular partners, such as DDB1, STAT proteins and, possibly, nuclear factors such as IRF3. The P protein binds to multiple partners as well, including N and the L protein. Therefore, intrinsic disorder represents an ergonomic solution for the virus to encode fewer proteins with more functions while keeping the genome size as small as possible. It is therefore not surprising to find large IDRs in proteins having a broad molecular partnership, such as the *Henipavirus* V proteins, and, more generally, in viral proteins [32, 33].

Intrinsic disorder also represents a strategy to alleviate evolutionary constraints on overlapping reading frames (such in the case of PNT that overlaps with the C protein) and to buffer the deleterious effect of mutations (i.e. IDRs/IDPs are more tolerant of substitutions compared to globular proteins) (for reviews see [33, 34]). Structural disorder is also known to allow uncoupling between affinity and specificity (i.e. low affinity coupled to decent specificity). However, the persistence of residual disorder in IDP complexes (in the form of

fuzzy appendages) and a partial preconfiguration of binding motifs prior to binding afford a way to attenuate the entropic penalty associated to the disorder-to-order transition thereby, ultimately, modulating the binding affinity. The involvement of IDRs in protein-protein interactions that need to be finely tuned provides an exquisite means to modulate the interaction strength: by tuning the extent of preconfiguration of the binding motifs and/or the length of neighboring fuzzy appendages, the virus can achieve an optimal binding strength. It is therefore conceivable that the presence within NiV PNT of long disordered regions flanking binding motifs, such as those involved in binding to N° or STAT1, may exert a role in regulating these critical interactions.

The present results constitute an asset for obtaining atomistic information in future interaction studies involving NiV PNT. Indeed, the availability of the chemical shifts of NiV PNT will be instrumental to map the residues involved in binding to partners such as DDB1 and/or STAT1 in the context of either the isolated PNT domain or the V protein. Once detailed information is available from interaction studies, inhibitors can be rationally conceived. Therefore, the present study can be regarded as the starting point towards the design of inhibitors abrogating the ability of this virus to escape the innate immune response. Given the high similarity between NiV and HeV PNT (56 %), it is conceivable that the results of these studies could be extrapolated to some extent to the HeV as well.

Materials and Methods

Protein expression and purification

The NiV PNT construct, encoding residues 1-406 of the NiV P protein with a C-terminal hexahistidine tag, has been already described [8]. Expression of unlabeled NiV PNT was carried out as previously described [8].

Isotopically labeled (either ^{15}N-^{13}C or ^{15}N) NiV PNT samples were prepared by growing at 37°C transformed *E. coli* T7 cells (New England Biolabs, Ipswich, MA, USA) bearing the pRARE plasmid (Novagen, Madison, WI, USA) in LB medium supplemented with 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} choramphenicol. When the OD_{600} reached 0.6, the culture was centrifuged at 4,000 rpm for 10 min and the pellet was resuspended in ¼ of the initial volume of M9 medium (6 g L^{-1} of Na_2HPO_4, 3 g L^{-1} of KH_2PO_4, 0.5 g L^{-1} of NaCl, 0.246 g L^{-1} of $MgSO_4$) supplemented with 1 g L^{-1} of $^{15}NH_4Cl$ and 2 g L^{-1} of either glucose or ^{13}C-glucose.

After one hour at 37 °C, IPTG was added to a final concentration of 0.5 mM, and the cells were subsequently grown at 37 °C for 3.5 h.

NiV PNT was purified as described in [8], except that 6 M GuHCl was added after the lysis step to both denature bacterial proteases and recover the recombinant protein also from inclusion bodies with the aim of improving protein stability and yield. After a 1 hour incubation at 4 °C, the sample was clarified and the supernatant was purified through immobilized metal affinity chromatography (IMAC). The fractions containing the recombinant protein were combined, and then loaded onto a Superdex 75 HR 16/60 column (GE, Healthcare). The elution buffer was 10 mM sodium phosphate pH 6.5, supplemented with 5 mM EDTA and 5 mM DTT. The fractions containing the protein were collected and conserved at -20°C.

SAXS studies

Prior to SAXS experiments, the sample was concentrated using 30 kDa Amicon Ultra Centrifugal Filters (Merk Milllipore, Darmstadt, Germany) and loaded onto a Superdex 75 HR 16/60 column using 20 mM Tris/HCl pH 8.0, 0.3 M NaCl, 5 mM DTT as elution buffer. All SAXS experiments were carried out at the ESRF on beamline BM29 at a working energy of 12.5 KeV. Data were collected on a Pilatus (1M) detector. The wavelength was 0.992 Å. The sample to detector distance of the X-rays was 2.847, leading to scattering vectors q ranging from 0.028 to 4.525 nm^{-1}. The scattering vector is defined as $q = 4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The exposure time was optimized to reduce radiation damage. SAXS data were collected at 20 °C using purified protein samples (50 μL each). Proteins concentrations were 0.5 and 1.2 g/L. Samples were loaded in a fully-automated sample charger. Ten exposures of 10 s each were made for each protein concentration and data were combined to give the average scattering curve for each measurement. Data points affected by aggregation, possibly induced by radiation damages, were excluded.

Data were analyzed using the ATSAS program package [35]. Data reduction was performed using the established procedure available at BM29, and buffer background runs were subtracted from sample runs. The R_g and forward intensity at zero angle I(0) were determined according to the Guinier approximation at low q values, in a qR_g range up to 1.3:

$$Ln[I(q)] = Ln[I_0] - \frac{q^2 R_g^2}{3}$$ (Eq. 1)

The forward scattering intensities were calibrated using water as references. The R_g and pair distance distribution function, P(r), were calculated with the program GNOM [36].

The theoretical value of R_g (in Å) of a protein composed by N amino acids was calculated from the following equations describing the behavior of a set of natively folded (R_g^{NF}) [37], chemically denatured (R_g^U) [16] and intrinsically disordered (R_g^{IDP}) [37] proteins of known length:

$$R_g^{NF} = \sqrt{(3/5)}4.75N^{0.29} \qquad \text{(Eq. 2)}$$

$$Log(R_g^U) = 0.58Log(N) + 0.40 \qquad \text{(Eq. 3)}$$

$$R_g^{IDP} = R_0 N^v \qquad \text{(Eq. 4)}$$

where R_0 is 2.54 ± 0.01 and v is 0.522 ± 0.001.

NMR spectroscopy experiments

Immediately before NMR studies, the NiV PNT sample was thawed, centrifuged at 12,000 rpm to remove any possible protein aggregate and then concentrated using 30 kDa Amicon Ultra Centrifugal Filters (Merk Milllipore, Darmstadt, Germany). All NMR spectra were acquired using ^{15}N-^{13}C isotopically labeled NiV PNT about 200 μM.

The 2D ^{1}H-^{15}N HSQC NMR spectra were acquired at 288 K with a 23.5 T Bruker Avance III 950 NMR spectrometer. To accomplish the assignment of the protein, a series of triple-resonance 3/4/5D NMR experiments, specifically designed for IDPs and exploiting either ^{1}HN or ^{13}C detection, were acquired [38] (**Supplementary Table S1**). The projection reconstruction spectroscopy (APSY) approach was exploited in most of the multidimensional experiments to reduce the experimental time while preserving high spectral resolution in the indirect dimensions [39]. The 3D BEST-TROSY (BT) triple-resonance experiments [40] HN(CO)CACB, HNCACB, HNCANNH, HNCO, HN(CA)CO and HNCOCANNH were acquired to obtain the complementary information needed to complete the assignment, augmented by a 5D BT-HN(COCAN)CONH experiment [41] used to resolve a few ambiguities and to confirm the chemical shift assignment obtained with the 3D spectra. The ^{13}C-detected 3D CBCACON, CBCANCO [42], (H)COCON [43], and the 5D (HCA)CONCACON spectra [44, 45] were acquired at 288K with a 16.4 T Bruker AVANCE NEO 700 spectrometer equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. These

experiments provided additional information necessary to complete the sequence specific assignment.

All the heteronuclear relaxation experiments (R_1, R_2 and ^1H–^{15}N NOEs) were acquired using ^{15}N labeled NiV PNT at about 100 μM. The spectra were recorded at 288 K on a 16.4 T Bruker Avance NEO 700 spectrometer equipped with a cryogenically cooled triple resonance probehead. The ^{15}N R_1 and R_2 experiments were acquired with 8 scans (2048 x 256 points) and a relaxation delay of 3.0 s. To determine the ^{15}N R_1 the following delays were used: 20 ms, 60 ms, 120 ms, 180 ms, 250 ms, 400 ms, 500 ms, 600 ms, 750 ms, 900 ms. To determine the ^{15}N R_2 the following delays were used: 32 ms, 64 ms, 96 ms , 128 ms, 160 ms, 190 ms, 260 ms, 320 ms, 380 ms, 440 ms, 500 ms. The relaxation delay for ^{15}N R_1 and R_2 experiments was 3.0 s. The ^1H–^{15}N NOEs experiments were acquired with 64 scans (2048 x 288 points) and a relaxation delay of 6.0 s. The CLEANEX experiments were acquired with 16 scans (2048 x 248 points) and a relaxation delay of 3.0 s. The following delays were used: 5 ms, 10 ms, 20 ms, 30 ms.

NMR data processing and analysis

NMR data sets were processed using either Bruker TopSpin 3.5pl7 or TopSpin 4.0.6 software. CARA [46] was used to analyze and annotate the spectra.

The ^{15}N relaxation rates (R_1 and R_2) were determined by fitting the cross-peak intensity measured as a function of variable delay, to single-exponential decay using the Bruker Dynamic Center 2.4, available as stand-alone ancillary software of TopSpin by Bruker. H-^{15}N NOE values were obtained as a ratio between peak intensity in spectra recorded with and without ^1H saturation.

The secondary structure propensity (ncSSP) [22,23] from heteronuclear chemical shifts was determined by using the neighbor corrected structural propensity calculator (ncSPC) tool [23] available online at http://nmr.chem.rug.nl. The Mulder random coil chemical shift library [24] was chosen for the analysis.

Generation of conformational ensembles

To obtain the ensemble description of NiV PNT, various conformational ensembles were generated using different sets of experimental data. For the generation of SAXS-based

ensembles, we generated the initial pool of conformers using either RANCH [28] or Flexible-Meccano [47] and then selected the final sub-ensemble that best fits the experimental SAXS data using GAJOE [28]. While RANCH generates Cα-only conformers, Flexible-Meccano generates backbone-atoms conformers. Irrespective of the procedure used, the initial ensemble consisted of 10,000 conformers. The programs were used with default parameters.

To generate NMR-based ensembles, we first generated the initial pool of 10,000 NiV PNT conformers using Flexible-Meccano supplementing the program with the ncSSP data obtained as reported above. Due to technical limitations, only ncSSP values exceeding ± 0.1 were taken into account. Initial ensembles were also generated by imposing an α-helical conformation to residues 1-30. The various pools were then restricted using the program GAJOE using the default parameters, unless differently specified.

References

[1] Wang LF, Yu M, Hansson E, Pritchard LI, Shiell B, Michalski WP, et al. The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. J Virol. 2000;74:9972-9.

[2] Marsh GA, de Jong C, Barr JA, Tachedjian M, Smith C, Middleton D, et al. Cedar virus: a novel Henipavirus isolated from Australian bats. PLoS Pathog. 2012;8:e1002836.

[3] Bloyet LM, Welsch J, Enchery F, Mathieu C, de Breyne S, Horvat B, et al. HSP90 Chaperoning in Addition to Phosphoprotein Required for Folding but Not for Supporting Enzymatic Activities of Measles and Nipah Virus L Polymerases. J Virol. 2016;90:6642-56.

[4] Yabukarski F, Lawrence P, Tarbouriech N, Bourhis JM, Delaforge E, Jensen MR, et al. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. Nat Struct Mol Biol. 2014;21:754-9.

[5] Park MS, Shaw ML, Munoz-Jordan J, Cros JF, Nakaya T, Bouvier N, et al. Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. J Virol. 2003;77:1501-11.

[6] Fontana JM, Bankamp B, Rota PA. Inhibition of interferon induction and signaling by paramyxoviruses. Immunol Rev. 2008;225:46-67.

[7] Shaw ML, Garcia-Sastre A, Palese P, Basler CF. Nipah virus V and W proteins have a common STAT1-binding domain yet inhibit STAT1 activation from the cytoplasmic and nuclear compartments, respectively. J Virol. 2004;78:5633-41.

[8] Habchi J, Mamelli L, Darbon H, Longhi S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. PLoS ONE. 2010;5:e11684.

[9] Blocquel D, Beltrandi M, Erales J, Barbier P, Longhi S. Biochemical and structural studies of the oligomerization domain of the Nipah virus phosphoprotein: Evidence for an elongated coiled-coil homotrimer. Virology. 2013;446:162-72.

[10] Bruhn-Johannsen JF, Barnett K, Bibby J, Thomas J, Keegan R, Rigden D, et al. Crystal structure of the Nipah virus phosphoprotein tetramerization domain. J Virol. 2014;88:758-62.

[11] Habchi J, Blangy S, Mamelli L, Ringkjobing Jensen M, Blackledge M, Darbon H, et al. Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. J Biol Chem. 2011;286:13583-602.

[12] Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol. 2003;84:3239-52.

[13] Li T, Chen X, Garbutt KC, Zhou P, Zheng N. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. Cell. 2006;124:105-17.

[14] Salladini E, Delauzun V, Longhi S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. Mol Biosyst. 2017;13:2254-67.

[15] Karlin D, Belshaw R. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. PLoS ONE. 2012;7:e31719.

[16] Bernado P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J. 2009;97:2839-45.

[17] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110:13392-7.

[18] Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102-12.

[19] Das RK, Huang Y, Phillips AH, Kriwacki RW, Pappu RV. Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. Proc Natl Acad Sci U S A. 2016;113:5616-21.

[20] Sherry KP, Das RK, Pappu RV, Barrick D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proc Natl Acad Sci U S A. 2017;114:E9243-E52.

[21] Tedeschi G, Salladini E, Santambrogio C, Grandori R, Longhi S, Brocca S. Conformational response to charge clustering in synthetic intrinsically disordered proteins. Biochim Biophys Acta Gen Subj. 2018;1862:2204-14.

[22] Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Protein Sci. 2006;15:2795-804.

[23] Tamiola K, Mulder FA. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem Soc Trans. 2012;40:1014-20.

[24] Tamiola K, Acar B, Mulder FA. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc. 2010;132:18000-3.

[25] Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. Biochemistry. 2012;51:2224-31.

[26] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292:195-202.

[27] Milles S, Jensen MR, Lazert C, Guseva S, Ivashchenko S, Communie G, et al. An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. Sci Adv. 2018;4:eaat7778.

[28] Tria G, Mertens HDT, Kachala M, Svergun D. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ. 2015;2:202-17.

[29] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. FEBS J. 2005;272:5129-48.

[30] Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit. 2005;18:343-84.

[31] Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol. 2006;2:e100.

[32] Uversky VN, Longhi S. Flexible viruses: structural disorder in viral proteins. Hoboken, New Jersey: Joh Wiley & Sons; 2012.

[33] Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, et al. Structural Disorder in Viral Proteins. Chem Rev. 2014;114:6880-911.

[34] Longhi S, Bloyet LM, Gianni S, Gerlier D. How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. Cell Mol Life Sci. 2017;74:3091-118.

[35] Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. . J Appl Cryst. 2009;42:342-6.

[36] Svergun D. Determination of the regularization parameters in indirect-trasform methods using perceptual criteria. J Appl Cryst. 1992;25:495-503.

[37] Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry. 1999;38:16424-31.

[38] Brutscher B, Felli IC, Gil-Caballero S, Hosek T, Kummerle R, Piai A, et al. NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. Adv Exp Med Biol. 2015;870:49-122.

[39] Hiller S, Fiorito F, Wuthrich K, Wider G. Automated projection spectroscopy (APSY). Proc Natl Acad Sci U S A. 2005;102:10876-81.

[40] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR. 2013;55:311-21.

[41] Piai A, Hosek T, Gonnelli L, Zawadzka-Kazimierczuk A, Kozminski W, Brutscher B, et al. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. J Biomol NMR. 2014;60:209-18.

[42] Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson. 2009;198:275-81.

[43] Mateos B, Konrat R, Pierattelli R, Felli IC. NMR Characterization of Long-Range Contacts in Intrinsically Disordered Proteins from Paramagnetic Relaxation Enhancement in (13) C Direct-Detection Experiments. ChemBioChem. 2019;20:335-9.

[44] Bermel W, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, et al. High-dimensionality 13C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR. 2013;57:353-61.

[45] Murrali MG, Schiavina M, Sainati V, Bermel W, Pierattelli R, Felli IC. (13)C APSY-NMR for sequential assignment of intrinsically disordered proteins. J Biomol NMR. 2018;70:167-75.

[46] Keller R. The computer aided resonance assignment tutorial. Goldau: Cantina Verlag; 2004.

[47] Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics. 2012;28:1463-70.

[48] DeLano WL. The PyMOL molecular graphics system Proteins: Structure, Function and Bioinformatics. 2002;30:442-54.

[49] Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem. 1994;18:269-85.

[50] Cho MK, Kim HY, Bernado P, Fernandez CO, Blackledge M, Zweckstetter M. Amino acid bulkiness defines the local conformations and dynamics of natively unfolded alpha-synuclein and tau. J Am Chem Soc. 2007;129:3032-3.

[51] Svergun DI, Baraberato C, Koch MH. CRYSOL - a Program to evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Cryst. 1995;28:768-73.

[52] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605-12.
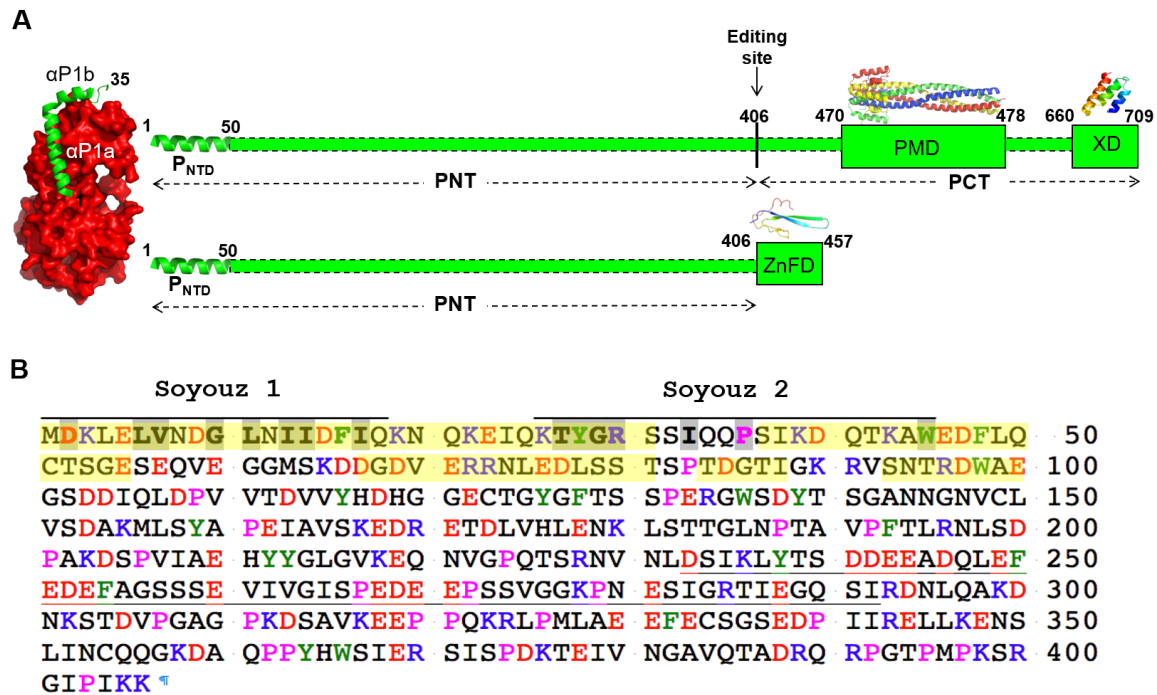
Figures

Figure 1. Modular organization and amino acid sequence of NiV PNT. (A) Modular organization of P. Domain organization of P showing two moieties, PNT and PCT. The P editing site is shown. Structured and disordered regions are represented as large and narrow boxes respectively. PNT: N-terminal region of P; PCT: C-terminal region of P. PMD: P multimerization domain; XD: X domain consisting of a triple α-helical bundle. The α-MoRE at the N-terminal region of P (P_{NTD}), which is partly preconfigured in solution and shown to adopt a stable α-helical conformation upon binding to the monomeric form of N [4], is shown as a green helix. The crystal structures of NiV PMD (PDB code 4N5B) [10] and of the NiV N°-P_{NTD} complex (PDB code 4CO6) are shown. The homology-derived models of NiV XD [11] and of the ZnFD of V are also shown [14]. All structures were drawn using Pymol [48]. **(B)** Amino acid sequence of NiV PNT (Uniprot code Q9IK91). Basic and acidic residues are shown in blue and red, respectively. Aromatic residues are shown in green and prolines in pink. The soyuz1 and soyuz2 motifs conserved in *Paramyxoviridae* members are shown. The low sequence complexity region (as obtained using SEG [49] with a window size of 25 residues, trigger segment complexity of 3.0 and extension segment complexity of 3.3) is underlined. Fluctuating α-helices α-P1 to α-P5, as observed in NMR studies of P_{1-100} [4] are shown in yellow.
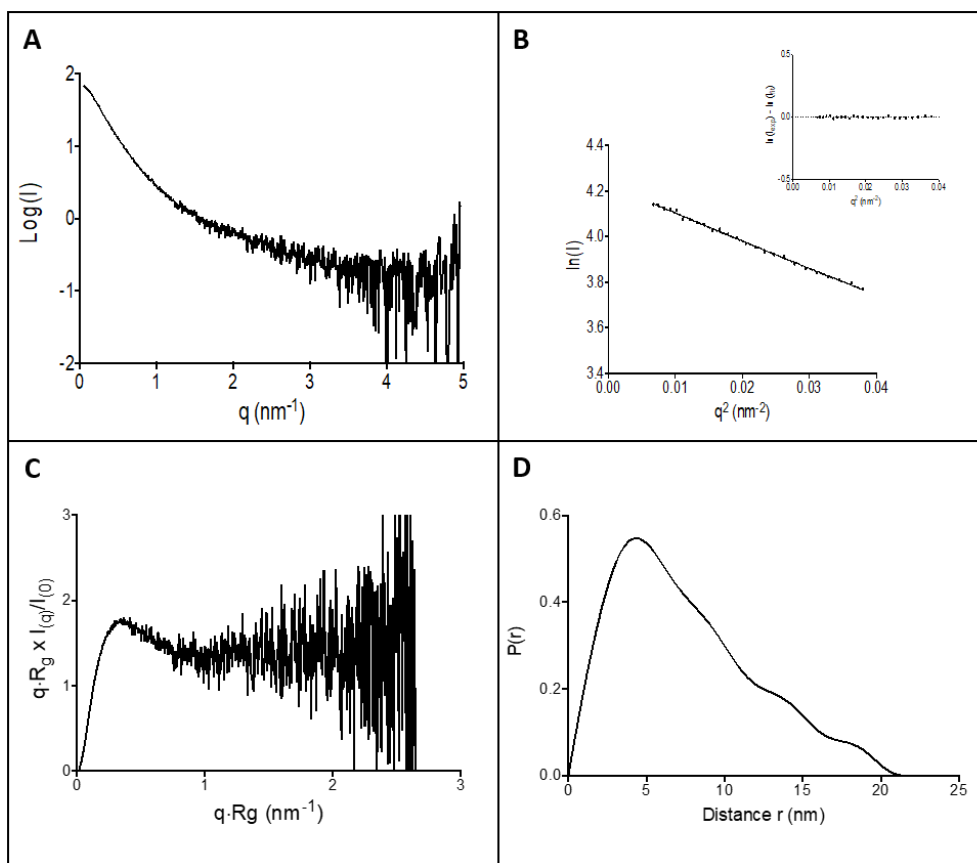
Figure 2. **SAXS studies of NiV PNT.** (**A**) SAXS scattering curve obtained at 1.5 g/L. (**B**) Guinier plot of the scattering curve at 1.5 g/L. Inset: residuals. (**C**) Pair distance distribution, P(r), function of the scattering data at 1.5 g/L. (**D**) Kratky plot of the scattering data at 1.5 g/L.
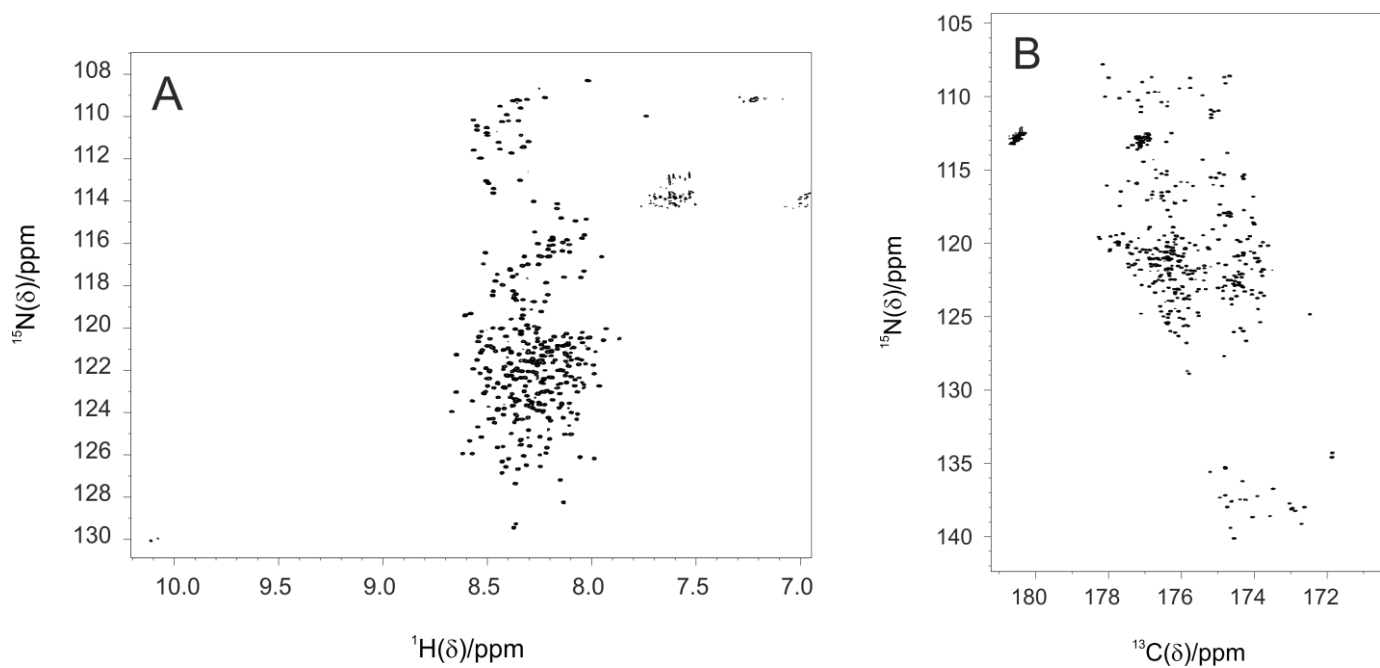
Figure 3. NiV PNT NMR spectra. (**A**) ^{1}H-^{15}N HSQC and (**B**) ^{13}C'-^{15}N CON spectra of NiV PNT. The spectra of ^{13}C-^{15}N-labeled NiV PNT (200 μM in 10 mM sodium phosphate buffer at pH 6.5) were recorded at 288 K due to sample instability at higher T.
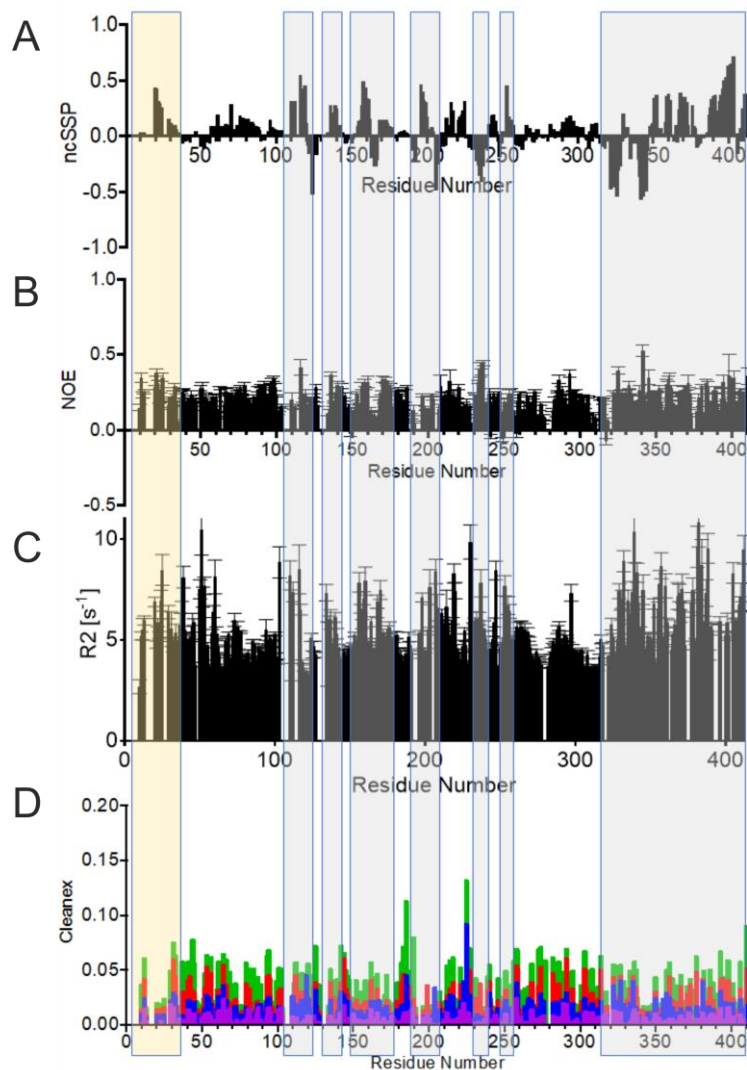
Figure 4. Characterization of the structure and dynamics of NiV PNT. (**A**) Neighbor-corrected secondary structure propensity (ncSSP) of PNT obtained from experimental HN, N, C', C^{α} and C^{β} chemical shifts. Positive and negative values correspond to α-helical and β-strand propensities, respectively. The regions highlighted in grey correspond to transiently structured elements. The region highlighted in yellow (aa 1-30) corresponds to helices αP1a and αP1b as observed in the crystal structure of the NiV N°-P_{NTD} complex (PDB code 4CO6). (**B**) ^{1}H-^{15}N steady state NOEs of PNT. (**C**) ^{15}N R2 relaxation rates and bulkiness calculated as described in [50]. (**D**) Intensities of cross peaks as obtained in CLEANEX experiments with different mixing times ranging from 5 (violet), 10 (blue), 20 (red) and 30 (green) seconds. Measurements were carried out at 700 MHz and 288 K on a 200 μM ^{15}N-labeled sample.
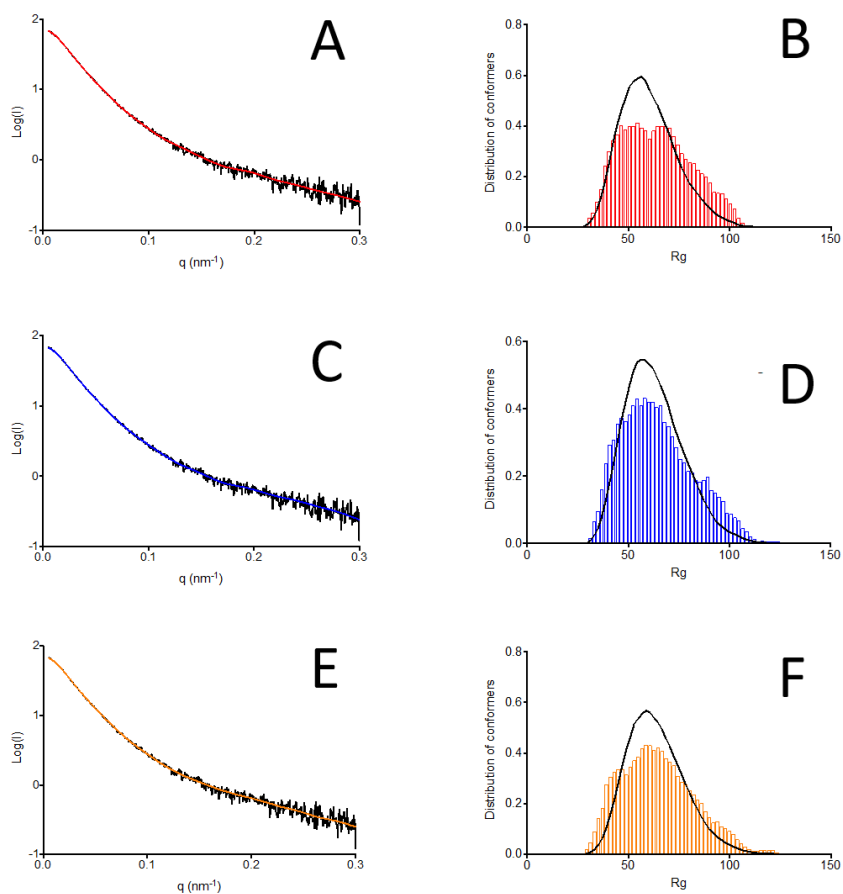
Figure 5. **Modeling NiV PNT as a conformational ensemble.** (**A, C, E**) Experimental scattering curve of NiV PNT at 1.5 g/L (black) and GAJOE fit as obtained using Crysol [51]. (**B, D, F**) R_g distribution of the initial ensemble generated by Flexible-Meccano (black) and of the selected sub-ensemble as obtained using GAJOE (colored bars). The initial pools of conformers were generated using Flexible-Meccano without (**B**) or with (**D, F**) NMR secondary structure restraints. In panels **F**, an α-helical conformation encompassing residues 1-30 was imposed during the generation of the ensemble by Flexible-Meccano.

Supplementary Material

Supplementary Table S1. Acquisition parameters for the NMR experiments described in the Material and Methods section.

Experiments	Dimension of acquired data			Spectral width (ppm)			n^a	d^b
	t_1	t_2	t_3	F_1	F_2	F_3		
^{1}H detected								
^{1}H – ^{15}N HSQC	1024 (^{15}N)	8192 (^{1}H)		35	15		2	0.03
BT HNCO	224 (^{13}C)	112 (^{15}N)	4096 (^{1}H)	6	24	12	4	0.13
BT HN(CA)CO	224 (^{13}C)	112 (^{15}N)	4096 (^{1}H)	6	24	12	8	0.13
BT HN(CO)CACB	220 (^{13}C)	128 (^{15}N)	4096 (^{1}H)	65	25	12	4	0.20
BT HNCACB	148 (^{13}C)	128 (^{15}N)	4096 (^{1}H)	34	25	12	16	0.20
BT HNCOCANNH	134 (^{13}C)	134 (^{15}N)	4096 (^{1}H)	25	25	12	8	0.20
BT HNCANNH	148 (^{13}C)	128 (^{15}N)	4096 (^{1}H)	25	25	12	16	0.20
^{13}C detected								
CON	1264 (^{15}N)	1024 (^{13}C)		40	30		8	2.00
CBCACON	164 (^{13}C)	128 (^{15}N)	1024 (^{13}C)	69	34	30	4	0.80
CBCANCO	164 (^{13}C)	128 (^{15}N)	1024 (^{13}C)	69	34	30	8	0.80
(H)COCON	108 (^{13}C)	128 (^{15}N)	1024 (^{13}C)	12	35	30	8	1.6

	t_1	t_2	t_3	t_4	t_5	F_1	F_2	F_3	F_4	F_5		
(HCA)CONCACON	192 (^{13}C)	192 (^{15}N)	192 (^{13}C)	192 (^{15}N)	2048 (^{13}C)	10	34	26	34	52	16	0.8

^{a} Number of scans
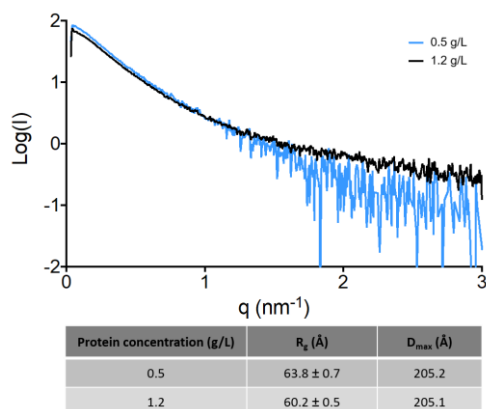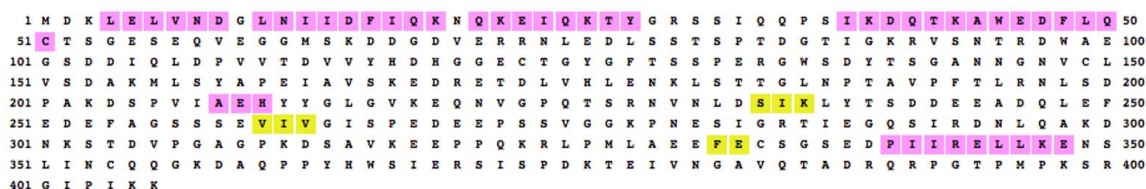^{b} relaxation delay in seconds

Figure S1. Experimental scattering curves of NiV PNT at 0.5 g/L (light blue) and 1.5 g/L (black). Shown are data for q values up to 3.0 nm^{-1} as obtained after correction for concentration.
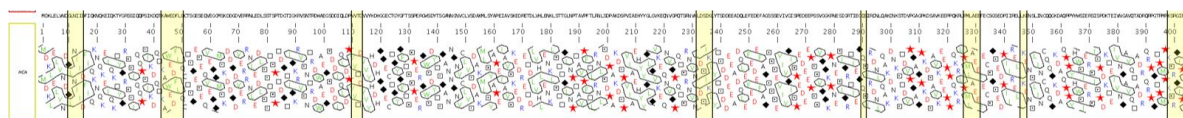
A

B

Figure S2. (**A**) Secondary structure elements as predicted by PSIPRED [26]. α-helices and β-strands are shown in pink and yellow, respectively. (**B**) Hydrophobic cluster analysis (HCA) plot [1] of NiV PNT as obtained using MeDor [2]. Molecular Recognitions Elements (MoREs), as predicted by MoRFPred [3] are shown in yellow.
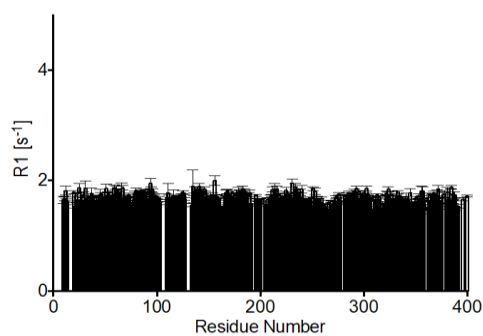

Figure S3. ^{15}N R1 relaxation rates of NiV PNT. Measurements were carried out at 700 MHz and 288 K on a 200 μM ^{15}N-labeled sample.

References

[1]   Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, et al. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci. 1997;53:621-45.

[2]   Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. BMC Genomics. 2008;9:S25.

[3]   Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, et al. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics. 2012;28:i75-83.

3.3 IDPs and liquid-liquid phase transitions

There is a growing interest in the topic of intracellular phase transitions that lead to the formation of biomolecular condensates. An increasing number of proteins have shown to undergo liquid-liquid phase separation in response to change in their environment. Interestingly, recent studies showed that the proteins that drive LLPS are often either IDPs of hybrid proteins with IDRs containing low complexity domains (LCDs) consisting of repeat amino acids with low diversity, favouring polar and charged groups.[102,103]

There are several reasons why the level of disorder may promote a phase transition.[104] Since IDPs typically contain a high number of charged residues, it is expected that electrostatic interactions would play an important role in conformational behaviour of IDPs. As a result, some IDPs have primary structures that are locally enriched in similarly charged residues, thereby containing regions of preferentially positively or negatively charged residues that might serve as good candidates for electrostatics-driven LLPS. Their lack of unique stable structures characterizes also the high sensitivity to changes in their environment. Thus, such environmental sensitivity and related capability to undergo fast environment-modulated transitions defines the role of IDPs in regulation of LLPS.

The discovery that proteins implicated in neurodegenerative aggregation such as Tau[105], α-synuclein[106] and huntingtin protein[107], can undergo liquid phase separation, rise the interest in this physical transition that often occur as preceding step of fibril formation. This phenomenon has been serendipity discovered also for the viral V protein, a protein from the Hendra virus introduced in Chapter 3.2 together with the corresponding protein from Nipah virus.

NMR spectroscopy demonstrated to be a useful complement to other biophysical techniques to unveil the mechanism of this physical transition. For example, in the following article NMR spectroscopy was combined with the negative-staining Transmission electron microscopy (TEM) to provide a bilateral estimation of the quantitative variation of the monomeric form of the protein. These results contribute to a large project on virus-induced cell toxicity carried out by Dr. Longhi's group at AFMB of Marseille.

3.3.1 Article: Phase transition and amyloid formation by a viral protein as an additional molecular mechanism of virus-induced cell toxicity

Phase transition and amyloid formation by a viral protein as an additional molecular mechanism of virus-induced cell toxicity

Edoardo Salladini[1], Claire Debarnot[1], Vincent Delauzun[1], Maria Grazia Murrali[2], Priscila Sutto-Ortiz[1], Silvia Spinelli[1], Roberta Pierattelli[2], Christophe Bignon[1] and Sonia Longhi[*]

[1]Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Marseille, France
[2]Magnetic Resonance Center (CERM), Department of Chemistry "Ugo Schiff" Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

*to whom correspondence should be sent
    Sonia Longhi
    AFMB, UMR 7257 CNRS and Aix-Marseille University
    163, avenue de Luminy, Case 932, 13288 Marseille Cedex 09, France
    Tel: (33) 4 91 82 55 80; Fax: (33) 4 91 26 67 20
    E-mail: Sonia.Longhi@afmb.univ-mrs.fr

Abstract

Henipaviruses are severe human pathogens responsible for severe encephalitis. Their V protein is a key player in the evasion of the host innate immune response. We have previously reported a biophysical characterization of the *Henipavirus* V proteins and shown that they interact with DDB1, a cellular protein that is a component of the ubiquitin ligase E3 complex. Here, we serendipitously discovered that the Hendra virus V protein undergoes a liquid-hydrogel phase transition. By combining experimental and bioinformatics approaches, we have identified the V region responsible for this phenomenon. This region (referred to as PNT3), which falls within the long intrinsically disordered region of V, was further investigated using a combination of biophysical and structural approaches. ThioflavinT and Congo red binding assays, together with negative-staining electron microscopy studies, show that this region forms amyloid-like, β-enriched structures. Such structures are also formed in mammal cells transfected to express PNT3. Those cells also exhibit a reduced viability in the presence of a stress agent. Interestingly, mammal cells expressing a rationally designed, non-amyloidogenic PNT3 variant (PNT3^{3A}), appear to be much less sensitive to the stress agent, thus enabling the establishment of a link between fibril formation and cell toxicity. The present findings therefore pinpoint a so far never reported possible mechanism of virus-induced cell toxicity.

Key words. Hendra virus; V protein, intrinsically disordered proteins, amyloids, fibrils, phase transitions.

Introduction

The Hendra and Nipah viruses (HeV and NiV) are members of the *Paramyxoviridae* family within the *Mononegavirales* order that comprises non-segmented, negative-stranded RNA viruses. NiV and HeV are zoonotic agents responsible for severe encephalitis in humans that have been classified within the *Henipavirus* genus [1]. HeV emerged in 1994 as the etiologic agent of a sudden outbreak of acute respiratory and neurological disease in horses in Brisbane, Australia. HeV infections presently constitute a serious threat to livestock in Australia, where sporadic and lethal transmission to humans has also occurred. After the first cases of human infection in 1998 in Malaysia, NiV has regularly reemerged since 2001 in Bangladesh, with an average case fatality of 70%. The ability of NiV to be transmitted by direct inter-human transmission further extends its potential to cause deadly outbreaks [2-4]. In addition, the discovery of Henipaviruses in other species of bats in West Africa and China [5] underscores the threat that these viruses constitute to human health. The susceptibility of humans, their high pathogenicity, the wide host range and interspecies transmission and the lack of vaccines and therapeutic treatments led to the classification of henipaviruses as biosecurity level 4 (BSL-4) pathogens and as potential bio-terrorism agents.

Like for all *Mononegavirales* members, the genome of henipaviruses is encapsidated by the nucleoprotein (N) within a helical nucleocapsid that serves as the substrate used by the viral polymerase for both transcription and replication. The viral polymerase is a complex made of the large (L) protein and the phosphoprotein (P). The P protein is an essential polymerase cofactor: not only it allows recruitment of L onto the nucleocapsid template but also serves as a chaperon for L [6, 7].

Henipavirus P proteins consist in an exceptionally long (> 400 aa), N-terminal disordered region (PNT) [8] and in a C-terminal region of approximately 300 aa containing two structured regions: a coiled-coil domain responsible for P multimerization (PMD) [9-11], and the X domain (XD). The latter adopts a triple α-helical bundle fold [12] and is responsible for interaction with the C-terminal disordered domain (N_{TAIL}) of the N protein [13-15]. In paramyxoviruses, the P gene also encodes the V and W proteins that are produced upon addition of either one (protein V) or two (protein W) non-templated guanosines at the editing site of the P messenger. The P, V and W proteins therefore share the PNT region that can be considered as a *bona fide* domain (**Figure 1A**). The C-terminal domain unique to paramyxoviral V proteins is predicted to be structured and to bind zinc [16] and is therefore referred to as zinc-finger domain (ZnFD).

Paramyxoviral V and W proteins are key players in the evasion of the interferon (IFN) mediated response. They act both by antagonizing IFN signaling and by inhibiting IFN induction [17, 18]. One of the IFN signaling pathways relies on STAT proteins (STAT1 and STAT2) activation and subsequent nuclear translocation. Once imported in the nucleus, they interact with IRF-9 to form the ISGF3 complex that activates the transcription of IFN-stimulated genes whose products inhibit viral replication [17]. The V and W proteins of henipaviruses have an antagonist activity of IFN signaling [17, 19]. They both bind to STAT1 via their PNT domain [20] leading to either inhibition of STAT1 translocation into the nucleus (V) or STAT1 sequestration in the nucleus (W) [20].

Although the *Henipavirus* P protein has an anti-IFN function too, its IFN antagonist activity is moderate compared to V (or W). This observation advocates for a critical role of the C-terminal domain of V and W in the anti-IFN function. That the *Henipavirus* V plays a pivotal role in the evasion of the innate immune response is corroborated by the fact that the Cedar virus (the lastly discovered *Henipavirus* member), which lacks the V protein, induces an IFN response much more pronounced compared to HeV, as well as an asymptomatic

infection in animal models [21]. The fact that NiV and HeV are the paramyxoviruses with the highest frequency of P messenger editing and also highly pathogenic comes in further support of a critical role of V in counteracting the innate immune response [18].

Several paramyxoviruses are able to hijack the cellular ubiquitin ligase E3 complex to promote the rapid degradation of STAT proteins. This activity relies on the ability of their V protein to bind to DNA damage-binding protein 1 (DDB1), a component of the ubiquitin ligase E3 complex, and then to recruit STAT proteins onto this complex [22]. In line with this, we have recently reported that *Henipavirus* V proteins interact with DDB1, and have unveiled a critical contribution of their C-terminal domain [23]. Incidentally, we also confirmed that the latter is folded and has a high β content both in isolation and when appended to PNT. Likewise, we have shown that PNT retains its overall disordered nature also in the context of the V protein. This finding rules out the possibility that PNT might adopt a unique conformation in the context of the V protein from which function could arise, and rather argues for a scenario where it is the C-terminal ZnFD that specifically confers to the V protein the ability to counteract the host immune response.

In the course of a further characterization of the *Henipavirus* V proteins, prompted by the fact that they are promising targets for antiviral approaches, we serendipitously discovered that the HeV V protein has the ability to form a hydrogel. In light of the growing number of studies pointing to a critical role of phase separations and transitions mediated by intrinsically disordered proteins (IDPs) and regions (IDRs) in various biological processes [24-26], we have decided to investigate in details this peculiar behavior.

By combining experimental and bioinformatics approaches, we have identified the V region responsible for this phenomenon and have further investigated it using a combination of biophysical and structural approaches. Using Congo Red (CR) and thioflavin T (ThT) binding assays and negative-staining electron microscopy we show that this region forms amyloid-like fibrils. Finally we show that mammal cells transfected to express this region display an increased sensitivity (i.e. increased mortality) towards a stress agent.

Results

Liquid-hydrogel transitions by the HeV V protein and identification of the region responsible for this behavior

In view of an in-depth structural characterization of the *Henipavirus* V proteins, we purified large amounts of both NiV and HeV V proteins at a relatively high concentration (i.e. ≥ 10 mg/mL, 200 μM) and stored them at -20°C. Upon thawing them, we noticed that the HeV V protein forms a hydrogel (**Figure 1B**). This liquid-hydrogel transition is irreversible as neither dilution nor boiling can restore the liquid state.

In order to identify the region responsible for this peculiar behavior, we resorted to dividing the V protein in various fragments. Out of the two already available HeV domains (i.e. ZnFD and PNT) [8, 23], only PNT was found to be able to form a hydrogel under conditions similar to those used for HeV V (**Figure 1C**). We subsequently divided the HeV PNT domain in four overlapping fragments (referred to as PNT1-PNT4) of 110 residues each (**Figure 1A**). The PNT fragments were all purified to homogeneity (**Figure 1D**). PNT3 (aa 200-310) was identified as the only V protein fragment able to form a gel after a freezing/thawing cycle at 200 μM (**Figure 1E**). This ability is also retained in the context of a PNT3-GFP fusion although gel formation was observed at a higher protein concentration (**Figure 1F**).

The ability of PNT3 to form a hydrogel after a freezing/thawing cycle is in line with bioinformatics analyses that unveiled the presence of a low complexity region (enriched in

4

Glu) encompassing residues 240-307. Low complexity domains are indeed known to drive physiologically reversible assembly of IDPs into membrane-free organelles, liquid droplets and hydrogel-like structures [26-30]. In addition, PNT3 contains a stretch of three contiguous tyrosines (aa 210-212 of P), and two non-contiguous tyrosines at positions 237 and 249 (**Figure 1A**). The well-established effect of tyrosines [31] and, more generally, of π-orbital containing residues [32, 33] in promoting phase separation provides an additional conceptual piece to rationalize the behavior of PNT3.
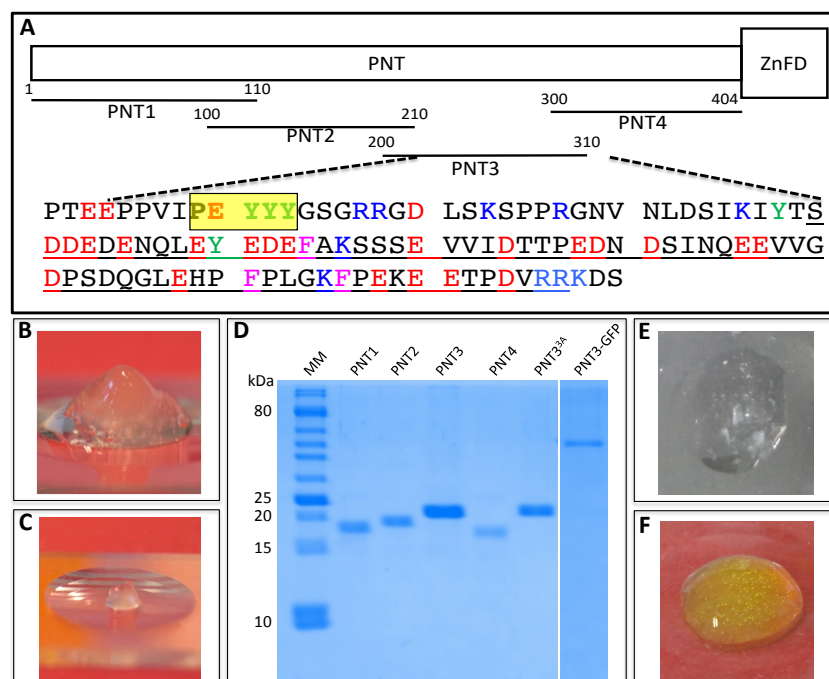

Figure 1. (**A**) Schematic organization of the HeV V protein. The intrinsically disordered P N-Terminal domain (PNT) and the zinc-finger domain (ZnFD) are represented as a narrow and large box respectively. The various PNT fragments herein generated are shown. The amino acid sequence of PNT3 is shown, with the most amyloidogenic region, as predicted by FoldAmyloid, framed in yellow. The low complexity region is underlined. Tyr residues are shown in green, acidic residues in red and basic residues in blue. (**B**) and (**C**) Hydrogels formed by concentrated (200 μM) V (**B**) and PNT (**C**) proteins upon freezing and thawing. (**D**) SDS-PAGE analysis of purified PNT fragments, of PNT3[3A] and of PNT3-GFP. MM: molecular mass markers. (**E**) and (**F**) Hydrogels formed by PNT3 at 200 μM (**E**) and PNT3-GFP at 1 mM (**F**) upon freezing and thawing.

Liquid-hydrogel transitions are thought to be preceded by a step of liquid-liquid phase separation (LLPS) [26, 34] that can be tackled by a wide range of physico-chemical solution studies. We thus first assessed whether PNT3 phase separates and then investigate the conditions under which this phenomenon occurs.

PNT3 undergoes phase separation

After a 1-hour incubation at room temperature (RT) in the presence of a crowding agent (i.e. 30% PEG[300]), PNT3 was found to phase separate in the 80 – 240 μM range and to form a hydrogel at 320 μM (see inset in **Figure 2A**). Phase separation can be quantified by turbidity measurements (**Figure 2A**) that showed that the phenomenon is dependent on both protein and PEG concentration, with a PEG concentration of 20% having no significant impact. We next assessed whether PNT3 retains the ability to aggregate in the absence of a crowding agent and investigated the impact of salt and temperature by monitoring the formation of aggregates. As shown in **Figure 2B**, aggregation occurs in a concentration-dependent manner even in the absence of PEG. Interestingly, salt does not seem to significantly affect the ability of the protein to form aggregates, suggesting that the

phenomenon is triggered by the formation of a hydrophobic core. The formation of aggregates is slightly enhanced at 37°C (**Figure 2B**).
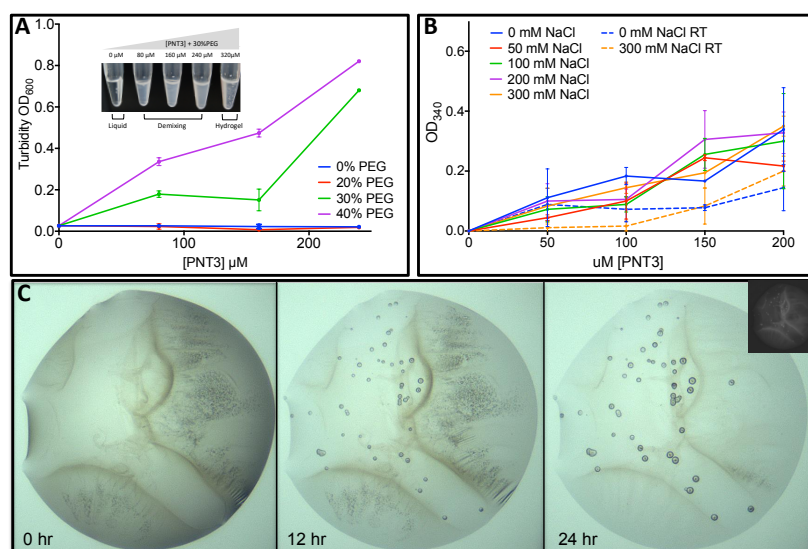

Figure 2. Liquid-liquid phase separation (LLPS) of PNT3. (A) Turbidity measurements of PNT3 samples at different concentrations either in the absence or in the presence of increasing PEG 400 concentrations after 1 hour incubation at RT. (**B**) Aggregate formation in PNT3 samples at different concentrations either in the absence or in the presence of increasing concentrations of NaCl after 1 hour incubation at 37°C (continuous lines) or at RT (dotted lines). (**C**) Droplets of 0.3 µl of 250 µM PNT3 in the presence of 10% PEG 4000 at day 0, 1 and 2 of incubation at RT.

A time course analysis of phase separated PNT3, as obtained after incubation at RT of a PNT3 solution at 250 µM supplemented with 10% PEG_{4000}, shows the appearance of droplets after 1 day of incubation (**Figure 2C**, central panel). Those droplets become larger between 24 and 48 hours, exhibit a relatively slow dynamics and seem to be rather stable (i.e. all droplets observed after one day can also be observed in the same position after two days). These features suggest a solid-like nature. Indeed, if the droplets were liquid like, Ostwald ripening (i.e. small droplets disappearing in favor of large ones) or coalescence would be expected. Definite answers as to the liquid *versus* solid nature of these droplets in this regard however await fluorescence recovery after photobleaching (FRAP) [35] studies. The proteinaceous nature of those droplets was confirmed under UV light (see inset in **Figure 2C**).

In conclusion, these experiments show that the phase separation undergone by PNT3 is multiparametric, depending on protein concentration, temperature, time and crowding agents.

PNT3 self-assembles into amyloid-like supramolecular structures

We next sought at shedding light onto the nature of these aggregates. To this end, we analyzed the ability of PNT3 to bind the amyloid-specific dye CR [36]. When the dye is bound to cross β-sheet structures there is hyperchromicity and a red shift of the absorbance maximum. The addition of 25 µM PNT3 promotes a shift in the CR spectrum from 497 nm to 515 nm (**Figure 3A**).

This observation provides the first hint suggesting that PNT3 can form β-enriched/amyloid-like structures. To quantify this phenomenon, we carried out binding assays with ThT, another well-known amyloid-specific dye [37]. Binding of ThT to PNT3 was found to induce a large enhancement in the intensity of the ThT fluorescence emission spectrum in a time- and concentration-dependent manner (**Figures 3B** and **3C**). In particular, while PNT3 samples at 1.5 and 2.5 µM give rise to a very similar kinetics, a significantly faster kinetics is

observed at 6.0 μM (**Figures 3B**). When the protein concentration is increased to 20 μM the kinetics is much faster, with the plateau being reached already after 2 hours (**Figures 3B**).
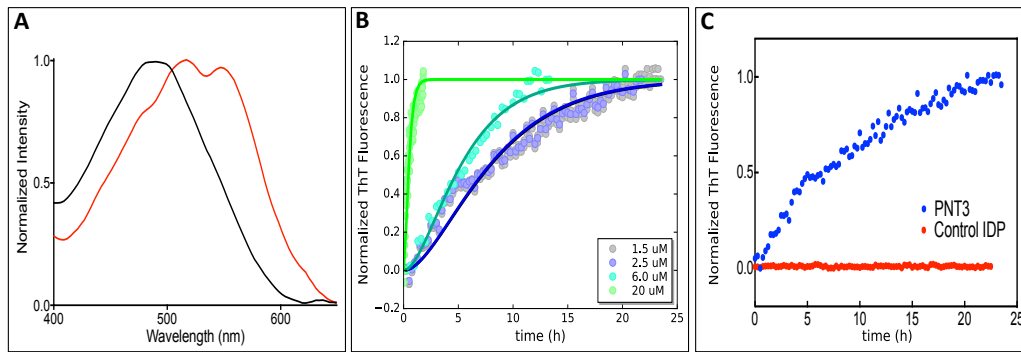

Figure 3. Congo Red and Thiofavin T binding assays. (A) Congo red (CR) spectral changes in the absence (black line) or in the presence (red line) of a 25 μM PNT3 sample at after 27 hours of incubation at 37°C. Note the characteristic shift from 497 nm to 515 nm typical of the dye when bound to amyloid-like aggregates. (**B**) Thioflavin T (ThT) binding assays of PNT3 at various concentrations and at various time points of incubation at 37°C. (**C**) ThT binding assays of PNT3 and of a control IDP (measles virus N_{TAIL}) at 6 μM. The fluorescence was measured at various time points of incubation at 37°C.

The ability of PNT3 to bind both CR and ThT is consistent with bioinformatics analyses carried out using the FoldAmyloid predictor (http://bioinfo.protres.ru/fold-amyloid/) that identified a motif (aa 208-212) containing three contiguous tyrosines as the most amyloidogenic region (see **Figure 1A**). This is also in line with the lack of a significant effect of salt on PNT3 aggregation (**Figure 2B**), indicative of the presence of a hydrophobic core. These results therefore argue for the ability of PNT3 to form amyloid-like assembles.

Far-UV circular dichroism studies of PNT3

To achieve additional insights on the secondary structure content of PNT3 before and after phase separation, we recorded the far-UV CD spectra of PNT3 under different conditions (**Figure 4A**). The CD spectrum of a freshly purified sample of PNT3 is typical of an IDP, as illustrated by the very pronounced negative peak at 200 nm. After incubation at 37 °C for ≥ 16 hours, i.e. a condition where we could document fibril formations in ThT and CR binding assays (see **Figure 3**), a dramatic decrease in the signal was observed. This phenomenon likely arises from the formation of fibrils that cannot be crossed by the polarized light, and was also observed in the case of phase separated Pro-Arg dipeptide repeats [29]. Interestingly, the CD spectrum of PNT3 hydrogel droplets as obtained following a freezing/thawing cycle shows a disorder to β transition, as judged from the appearance of a peak at 215 nm (**Figure 4A**). Collectively these data support the conclusion that PNT3 forms β-enriched fibrils in phase-separated and hydrogels droplets.

X-ray diffraction studies of PNT3

To assess the presence of cross-β properties in the sample after the ThT assay we employed X-ray diffraction. Fiber diffraction allows visualizing the distances between hydrogen-bonded β-strands (4.7-8 Å diffraction signal on the meridian of the pattern) and the distance that occur from association of the sheets (10-12 Å on the equator, depending on the size of side chains). If the fibers are not aligned on a plane, the diffraction of those distances arises in every direction and appears as a circular intensity, otherwise it is possible to discriminate between meridian and equator axes [38-40]. As shown in **Figure 4B**, the resulting diffraction pattern shows circular reflections at 4.8 Å. There is also a slightly asymmetrical diffraction pattern at 8.2 Å that indicates the equator and suggests a partial alignment of the fibres. These reflections are prototypic of cross-β structure [38-40].
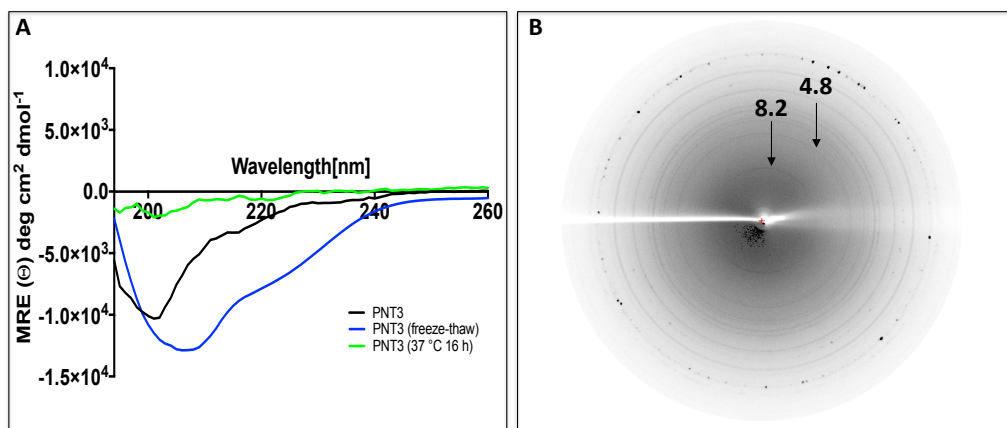
Figure 4. (A) Far-UV circular dichroism (CD) studies of PNT3. The spectra were recorded in 10 mM sodium phosphate pH 7 at 20°C. Protein concentration was 0.1 mg/mL (6 μM). The spectrum shown in black corresponds to a freshly purified PNT3 sample recorded immediately after elution from the SEC column. The spectrum shown in green corresponds to a sample incubated at 37°C over night. The spectrum shown in blue corresponds to a sample frozen and thawed. Note the shift from 200 nm to 215 nm of the negative peak in the gelified sample and the dramatic decrease in the signal for the sample incubated over night at 37°C. **(B)** X-ray fiber diffraction pattern of a polymerized PNT3 sample (see Materials and Methods). Prominent, circular X-ray reflections were observed at 4.8 Å and 8.2 Å, which is typical of cross-β structure. The spots at high resolution reflect the diffraction of salt crystals.

In combination with the ThT binding assays, the X-ray diffraction pattern provides strong evidence of the presence of amyloid-like polymers as the structural basis of phase-separated PNT3.

Small-angle X-ray scattering studies of PNT3

In view of achieving a better description of the evolution of the conformational properties of PNT3 over time, we carried out SAXS studies. Synchrotron SAXS data were collected from a PNT3 sample at two different concentrations and at different times of incubation (from 0 to 630 min) at 37°C (**Figure 5A** and **Supplementary Figure S1A**). Overall, the resulting scattering curves show an increase in intensities in the small-angle region with increasing incubation time (**Figure 5A** and **Supplementary Figure S1A**) consistent with an evolution from a monomeric to an oligomeric and/or fibrillar state. We followed this evolution using three parameters, namely the extrapolated forward scattering (I(0)), which is proportional to the molecular mass of the scatterer, the calculated radius of gyration (R_g) and the maximal internal dimension (D_{max}) (**Supplementary Table S1**).

Both D_{max} (**Figure 5B** and **Supplementary Figure S1B)** and R_g (**Supplementary Table S1)** increase with increasing incubation times though in a non-linear manner. In the case of the sample at 1 mg/mL, the monomeric species (calculated molecular mass of 16 kDa) persists until 120 min, while at 2 mg/mL the monomeric species (calculated molecular mass of 16 kDa) can no longer be detected after 90 min (**Supplementary Table S1**). The kinetics of formation of oligomers and/or fibrils is thus accelerated at higher protein concentrations, in agreement with results from ThT binding experiments (see **Figure 3C**). This concentration-dependence of fibril formation can be better appreciated in **Figure 5C**, where the R_g is plotted as a function of incubation time. The R_g value of the monomeric species (32.6 ± 1 Å) is in perfect agreement with the value expected for an IDP of the same size (32.6 Å) according to Flory's equation [41]. The calculated initial and final values of R_g, D_{max} and molecular mass (**Figure 5C** and **Supplementary Table S1**) are very similar at the two protein concentrations. The calculated molecular mass after 630 min (see **Supplementary Table S1**) is consistent with an oligomeric/fibrillar species made of ~100 PNT3 monomers.

For both protein concentrations, the Kratky plots reveal an overall gain of content in ordered structure with increasing time, as judged from the appearance of a progressively more

pronounced maximum (**Figure 5D** and **Supplementary Figure S1C**). This increase in ordered structure is consistent with the formation of fibrillar species.
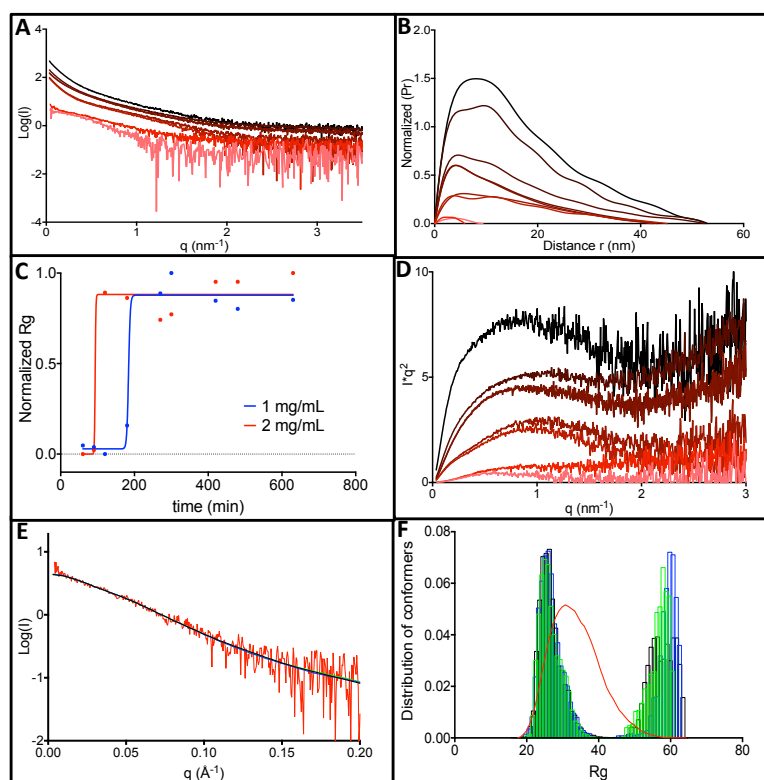

Figure 5. Small-angle X-ray scattering studies of PNT3. (**A**) Scattering curves as obtained from a sample of PNT3 at 2 mg mL^{-1} at various times of incubation at 37°C. The curves are represented with a color gradient ranging from pink to black with increasing incubation times. (**B**) Pair distance distribution, P(r), as obtained at 2 mg mL^{-1}. (**C**) Normalized R_g as a function of time for the two protein concentrations herein studied. Data were fitted to an asymmetric sigmoid using the Prism software. (**D**) Kratky plot of the SAXS data obtained at 2 mg mL^{-1}. The same color code as in panel A was used. (**E**) Experimental scattering curve of the protein at 2 mg mL^{-1} (red) and EOM fits as obtained by three independent runs (black, blue and green). (**F**) R_g distribution of the randomly generated conformers by EOM without constraints (red) and of the three sub-ensembles generated by three independent EOM runs (black, blue and green). Note the asymmetrical and bimodal R_g distribution.

We then investigated the distribution of conformations of the monomeric species at the highest protein concentration using the program suite EOM 2.0 [42]. From an initial pool of 10,000 random conformations, EOM selects a sub-ensemble of conformers that collectively reproduce the experimental SAXS data and represent the distribution of structures adopted by the protein in solution. The average SAXS scattering curves back-calculated from the selected sub-ensembles, as provided by three independent EOM runs, reproduce correctly the experimental curve (**Figure 5E**). The R_g distribution of the selected sub-ensembles, as obtained from the three independent EOM runs, are asymmetrical and bimodal, with two peaks centered at 25 Å and 60 Å (**Figure 5F**). The similarity of the R_g distributions obtained in different EOM runs attests of the reproducibility of the selection process and hence the reliability of the inferred conformational information.

These data indicate that the scattering curve of the monomeric species of PNT3 does not reflect a randomly distributed ensemble of conformations and R_g distributions and rather the existence of two distinct sub-populations differing in compaction. It is conceivable that the more compact form possesses a higher content in transiently populated secondary structures that imparts a higher propensity to undergo the disorder-to-order transition that accompanies the formation of β-enriched oligomers.

Nuclear magnetic resonance (NMR) and negative-staining transmission electron microscopy

(TEM) studies of PNT3

We next carried out negative-staining TEM and heteronuclear NMR studies to directly document fibril formation as well as any possible significant concomitant structural transition. We thus recorded the ^{1}H-^{15}N HSQC spectra of a uniformly labeled PNT3 sample after various incubation times at 37°C. As shown in **Figure 6**, a time-course analysis of the sample did not reveal any significant chemical shift displacement. Rather, an overall reduction in the peak intensities was observed with increasing incubation times (see also **Supplementary Figure S2**). A concomitant analysis with negative staining TEM unambiguously showed the progressive formation of amyloid-like fibrils. While those fibrils become progressively more abundant over time, their diameter (12-17 nm) seemingly remains unvaried. The observed reduction in peak intensities in NMR studies is akin to that observed in CD studies (see **Figure 4A**) and is attributable to the formation of fibrillar species that are not soluble and hence no longer contribute to solution-state NMR signals.
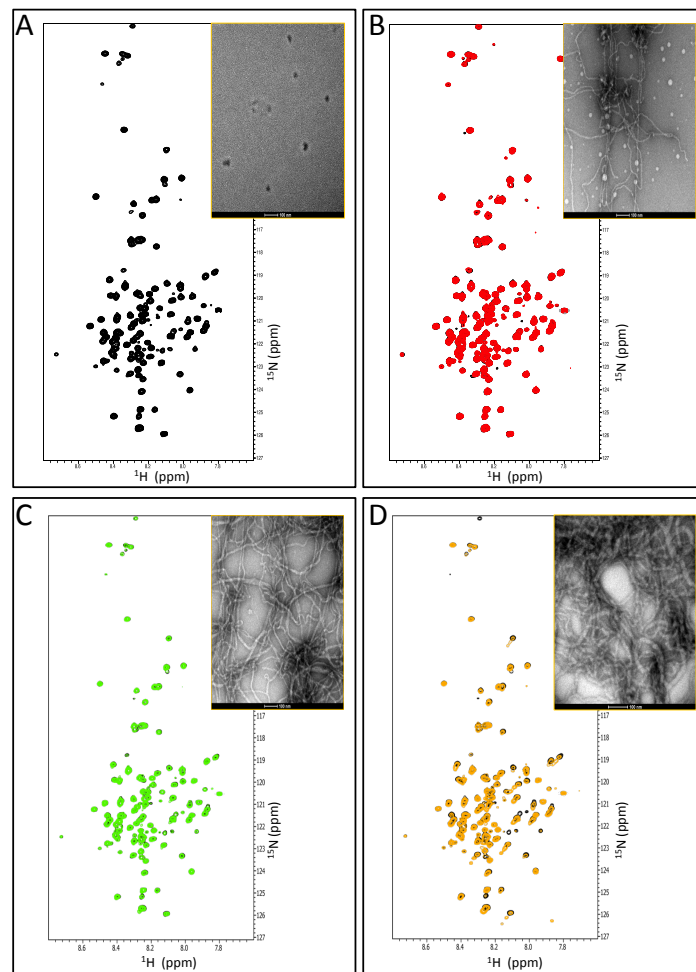

Figure 6. ^{1}H-^{15}N HSQC spectra of a PNT3 sample at 100 μM after 0 (**A**), 8 (**B**), 27 (**C**) and 56 (**D**) hours of incubation at 37°C. The HSQC spectrum of the sample at time 0 is shown in black in all panels. The spectra were recorded at 310 K. Insets: negative-staining TEM micrographs of a PNT3 sample at 200 μM incubated at 37°C for 0, 8, 27 and 56 hours.

Fragility of PNT3 amyloid-like fibers

Extreme stability is a hallmark of pathogenic amyloid fibers (see [43, 44] and references therein cited). In addition, yeast ultra-stable amyloids, derived from the low complexity sequences associated with transcription factors and RNA-binding proteins, have also been described [45]. These prion-like amyloid fibers share a common insensitivity to the

solubilizing effects of SDS. In order to investigate the SDS sensitivity of PNT3 fibers, heavily polymerized preparations of PNT3 (i.e. 100 μM after 57 hours of incubation at 37°C) were filtrated through a membrane allowing passage of monomeric proteins but not of fibers. As shown in **Figure 7A**, a very small amount of protein was found to pass through the filter when the fibers were diluted in standard buffer and filtrated immediately. By contrast, following incubation in the presence of 2% SDS at 37° C for 10 min, the amount of UV-adsorbing material passing into the filtrate increases. These results suggest that the fibers were at least partly depolymerized into monomers by SDS treatment. This was further supported by TEM studies that showed that in the presence of 5 mM SDS fibers disappear (**Figure 7B**).
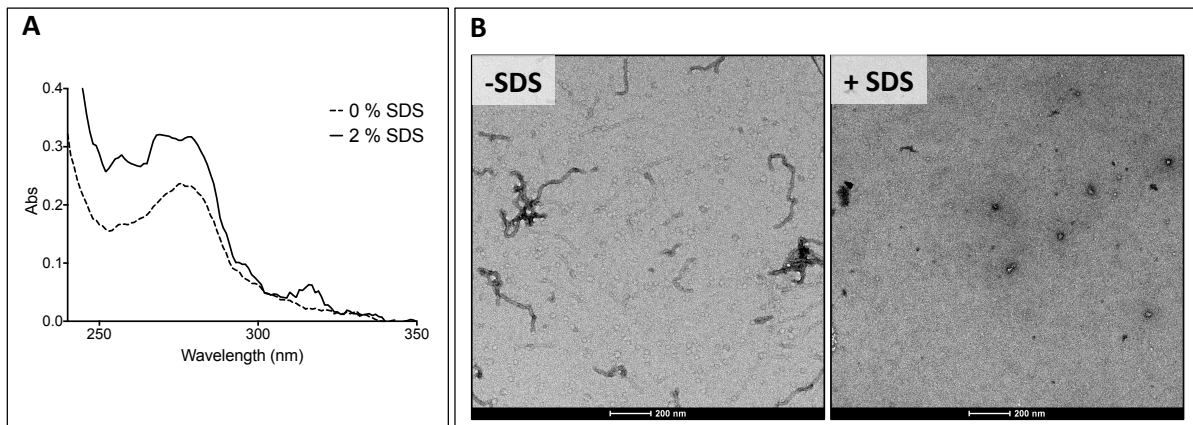

Figure 7. (**A**) UV adsorption of the filtrate of a sample of PNT3 at 100 μM after 56 hours of incubation at 37°C before (dashed line) and after (continuous line) exposure to 2% SDS and 10 min incubation at 37 C. Exposure of SDS enables more material to pass through the filter. (**B**) Micrographs of a PNT3 sample at 200 μM after 56 hours of incubation at 37°C either in the absence or in the presence of 5 mM SDS.

Indeed, analysis of up to 10 grid squares revealed the presence of as few as two small fibers all over. Thus, in the presence of SDS, PNT3 fibrils show a low stability, in line with previous observations on stress granule proteins [46], and on α-synuclein, tau and Aβ42 fibrils [47].

Therefore, although PNT3 fibers share morphological similarities and the presence of cross-β structure properties, they appear to be more fragile than the prion-like fibers broadly described in the literature.

The heat-shock protein 70 delays the fibrillation process of PNT3

In light of previous studies that documented the ability of chaperons, and in particular of the major inducible heat shock protein 70 (hsp70), to inhibit or delay fibril formation by prion-like proteins [48-52], we decided to ascertain whether human hsp70 has an impact on the fibrillation process of PNT3.

As shown in **Figure 8A**, ThT binding assays showed that hsp70 lowers the rate with which fibrillar species are formed. This observation is mirrored by TEM studies that showed that in the presence of hsp70 the formation of fibrils is hampered (**Figure 8B**).
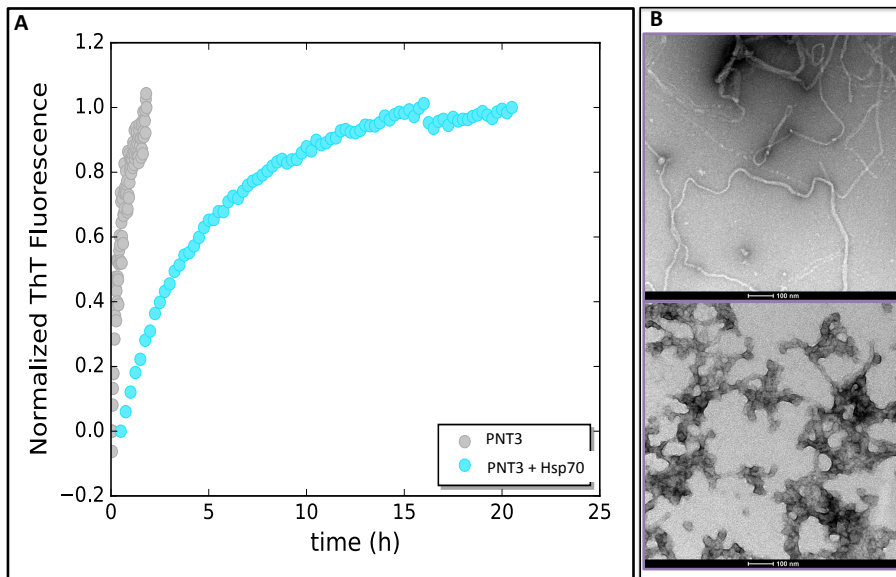
Figure 8. (A) ThT binding assays of a PNT3 sample at 20 μM either in the absence or in the presence of a two-fold molar excess of hsp70. The fluorescence was measured at various time points of incubation at 37°C. **(B)** TEM micrographs of a PNT3 sample at 100 μM incubated at 37°C for 27 hours either in the absence (top) or in the presence (bottom) of a two-fold molar excess of hsp70.

Rational design of a PNT3 variant with a hampered ability to form amyloid-like fibrils

In light of the well-documented role of Tyr residues in LLPS [31, 32], we reasoned that the YYY motif occurring at residues 210-212 of the V protein (see **Figure 1A**) might be responsible for the ability of PNT3 to phase separate and to form amyloid-like fibrils. We thus conceived and generated a PNT3 variant (referred to as PNT3[3A]) in which the three contiguous tyrosine residues were replaced with alanines. The purified PNT3[3A] variant (**Figure 1B**) has lost the ability to bind CR and ThT (**Figure 9A, B**). The inability of the variant to form amyloid-like fibrils was also confirmed by TEM studies (**Figure 9C**).
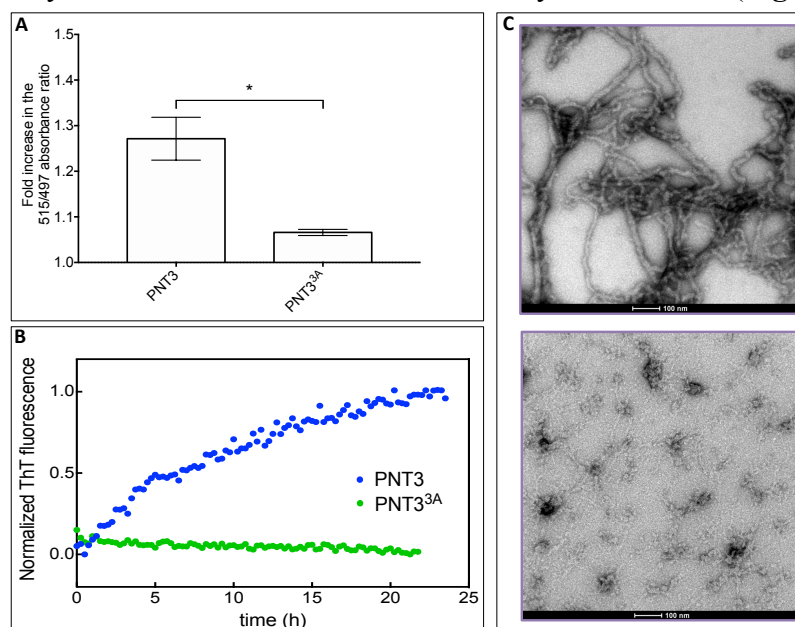

Figure 9. (A) CR binding by PNT3 and PNT3[3A]. Fold increase in the ratio between the absorbance at 515 and at 497 nm, with respect to a sample containing CR alone, of a PNT3 or a PNT3[3A] sample at 25 μM after 27 hours of incubation at 37°C. The error bar corresponds to the standard deviation, with n=3. The asterisk denotes a statistically significant difference ($p < 0.001$) (T Student's test). **(B)** ThT binding assays of PNT3 and of PNT3[3A] at 6 μM. The fluorescence was measured at various time points of incubation at 37°C. **(C)** TEM micrographs of PNT3 (top) and PNT3[3A] (bottom) at 200 μM and after a 56-hours incubation at 37°C.

12

Functional impact of PNT3 in mammal cells

In an attempt at unraveling the functional impact of the ability of PNT3 to phase separate, we carried out transfection experiments in mammal cells. CR staining of cells transfected to express PNT3 and PNT3-GFP showed that the proteins retain the ability to form amyloid-like fibrils also in the context of transfected cells (**Figure 10A**). Conversely, and as expected, CR fails to stain non-transfected cells (**Figure 10A**).

Next we sought at assessing any possible toxic effect related to the formation of amyloid-like fibrils by PNT3. To this end, we first measured cell viability in the presence of a stress agent (i.e. H_2O_2). As expected, transfected cells are more sensitive to a stress agent than non-transfected cells (**Figure 10B**). Interestingly, PNT3-GFP-transfected cells have a lower viability compared to GFP-transfected cells (**Figure 10B**). Notably, the expression level of PNT3-GFP is 20 times lower than that of GFP alone (as judged from fluorescence measurements, see **Supplementary Figure S3A**) thus ruling out the possibility that the lower toxicity of GFP may arise from a lower expression level. Results therefore advocate for a higher toxicity of PNT3-GFP, with the toxic effect of PNT3 being probably even underestimated.
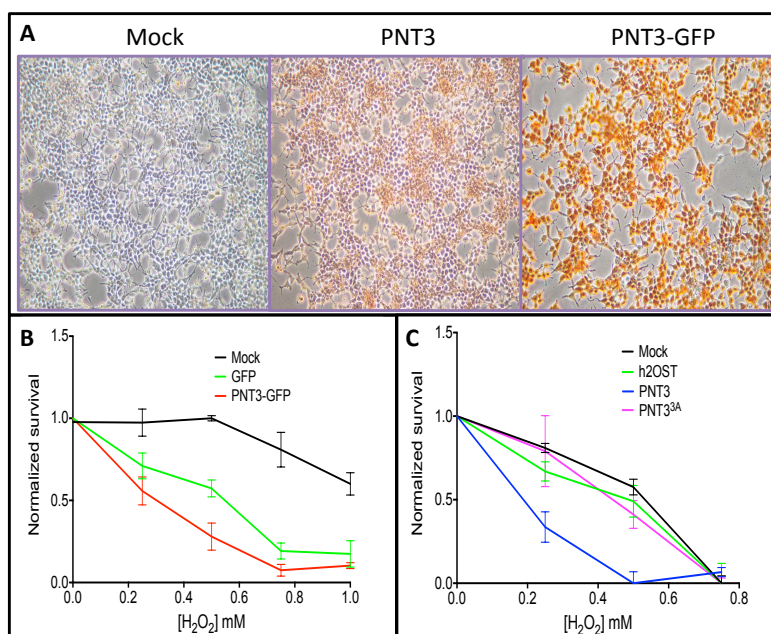

Figure 10. (A) CR staining of HEK cells non transfected or transfected to express PNT3 or PNT3-GFP. Note the red staining in the case of PNT3 and PNT3-GFP. **(B)** Viability assays of HEK cells non transfected or transfected to express GFP or PNT3-GFP after 6 hours of incubation at 37°C in the presence of increasing H_2O_2 concentrations. Shown are the mean values and s.d. as obtained from a triplicate. **(C)** Viability assays of HEK cells non transfected or transfected to express PNT3 or PNT3[3A] after 6 hours of incubation at 37°C in the presence of increasing H_2O_2 concentrations. Shown are the mean values and s.d. as obtained from a duplicate.

We next repeated these experiments with cells transfected to express either PNT3 or PNT3[3A], along with an irrelevant, control protein (i.e. 2OST). As shown in **Figure 10C**, the viability of cells expressing PNT3 in the presence of H_2O_2 is much lower compared to that of cells expressing PNT3[3A], the latter exhibiting a behavior close to that of non-transfected cells and of cells expressing the control protein. The possibility that the lack of toxicity of PNT3[3A] may arise from a much lower (or lack of) expression in transfected HEK cells was checked and ruled out (**Supplementary Figure S3B**).

Results advocate for a scenario where the substitution of the triple Tyr motif, which converts PNT3 into a non-amyloidogenic species (see **Figure 9**), also abrogates toxicity. These findings, especially if also corroborated by future CR staining experiments of PNT3[3A]-

13

transfected cells, allow establishing a link between cell toxicity and ability of the protein to aggregate into amyloid-like structures.

Discussion

In this paper we showed that the HeV V protein undergoes a liquid to hydrogel transition and mapped the region responsible for this behavior to the long intrinsically disordered region of V and, specifically, to residues 200-310 (PNT3). Using a combination of biophysical and structural approaches, we showed that *in vitro* PNT3 undergoes phase separation and forms amyloid-like, β-enriched structures. Noteworthy, such structures are also formed in mammal cells transfected to express PNT3. Those cells also exhibit a reduced viability in the presence of a stress agent. Interestingly, mammal cells expressing a rationally designed, non-amyloidogenic PNT3 variant (PNT3[3A]), appear to be much less sensitive to the stress agent, thus enabling the establishment of a link between fibril formation and cell toxicity.

In *Mononegavirales*, transcription and replication takes place in viral factories, e.g. cytoplasmic inclusions where the local concentration of viral proteins is increased thereby facilitating viral replication [53-58]. These viral factories, which can be either double-membrane [59] or membrane-less organelles [60], also likely prevent the activation of host innate immunity by shielding viral proteins and thus impairing their interaction with cellular antiviral proteins. It is conceivable that the ability of PNT3 to phase separate may be functionally coupled to the formation of viral factories. This however, would not explain the observed toxicity *per se* of PNT3 fibrils in mammal cells. Noteworthy, recent data suggest that phase transitions of proteins into liquid or hydrogel states could underlie pathological protein aggregation associated with neurodegenerative diseases. In particular, in the case of amyotrophic lateral sclerosis-associated protein FUS (fused in sarcoma) it has been proposed that the toxicity of irreversible FUS assemblies [61] may arise either from the fact that amyloid aggregates are intrinsically toxic for eukaryotic cells, or from the appearance of different density states in the cytoplasm that is the cause of a physiological dysfunction [62]. In this second scenario, the new membrane-less organelles could perturb cellular functions through either sequestration of particular proteins that are thus subtracted from their natural pathways or just by changing the dynamics inside the cells [63].

It is therefore tempting to speculate that the ability of HeV P/V proteins to form amyloids may constitute at least one of the possible molecular mechanisms underlying the pathogenicity of this virus and, possibly, also of its ability to cause encephalitis. The present findings therefore highlight a so far never reported possible mechanism of virus-induced cell toxicity. Definite conclusions however require additional experiments aimed at assessing the presence of amyloid-like fibrils in infected cells. Indeed, the possibility that amyloid formation in transfected cells may arise by a much higher expression level compared to that occurring in the natural context, cannot be ruled out.

To the best of our knowledge, the ability of a viral protein to form amyloids has been reported before only in the case of human papilloma virus (HPV)-16 E7 protein (i.e. the major oncoprotein of HPV). HPV E7 was indeed shown to be able self-assemble into defined spherical oligomers with amyloid-like properties [64, 65], although the possible functional implications of this phenomenon were only discussed in terms of the amyloid-cancer connection [66].

In light of the loss of ability to form amyloid-like structures by PNT3[3A], it is conceivable that the lower propensity of NiV V to undergo a liquid-hydrogel transition may reside in the presence of only two contiguous Tyr within the PNT3 region of NiV V instead of

three as in HeV V. Whether this has a functional relevance in terms of pathogenicity, immune response evasion and onset and progression of virus-induced encephalopathy remains to be established.

The relative fragility of PNT3 fibers, pinpointed by their sensitivity to SDS, might reflect a role as regulatory switches, i.e. fibers can form and disassemble in response to changes in the surrounding environment and this can play a role in the (in)activation of specific cellular pathways [25, 63, 67]. If fibers correspond to the solid-like structure within droplets resulting from phase separation, then their ability to disassemble might impart metastability (i.e. reversibility) to membrane-less organelles containing such amyloidogenic proteins. Given the link between material properties and pathological conditions [44, 61], cells have likely evolved mechanisms to monitor and control the fluidity of phase-separated droplets. Consistent with this, many ribonucleoprotein (RNP) bodies and granules are enriched in ATP-dependent chaperones such as hsp70 and hsp40 [68]. In agreement, the viscoelasticity of nucleoli and the dynamics of stress granule components exhibit a strong ATP dependence [69, 70]. Our results showing that hsp70 delays/impairs fibril formation by PNT3 are in line with these previous findings, and may reflect a mechanism whereby variations in hsp70 intracellular levels, typically occurring during viral infections (see [71] and references therein cited), may control the stability of phase separated droplets containing the P/V protein. In addition, the ability of hsp70 to hamper fibril formation by PNT3 may also be linked to the well-documented protective role of this chaperone against (paramyxo)viral infections [71-75].

Future efforts will focus on deciphering the underlying molecular mechanisms by which PNT3 fibrils reduces cell resistance to a stress agent in mammal cells. In particular, and in light of recent findings that showed that the Ebola virus VP35 protein (the functional counterpart of paramyxoviral P) blocks stress granule assembly, we will investigate possible interferences with stress granule (dis)assembly. Concomitantly, we will assess any possible synergistic contributions to the phenomenon of phase separation of PNT3 brought either by the C-terminal domain of the V and W proteins, or by oligomerization (i.e. in the context of the P protein).

Since the P protein is an essential component of the replicative complex of *Mononegavirales*, and since the presence of large IDRs is a wide spread and conserved property in *Mononegavirales* P proteins [16, 76-89], it is conceivable that the ability to undergo phase separation and transition is not unique to HeV P/V. As such, these results promise to have broad implications on a large number of important human pathogens.

Materials and Methods

Generation of the constructs

The constructs encoding *Henipavirus* V proteins, as well as their Zinc-finger and PNT domains, have already been described [8, 23].

The HeV PNT1, PNT2, PNT3 and PNT4 DNA fragments, encoding respectively aa 1-110, 100-210, 200-310 and 300-404 of the HeV P/V protein, were PCR-amplified using the the pDEST14/HeV PNT construct as template [8] and primers: PNT1-AttB1 and PNT1-AttB2; PNT2-AttB1 and PNT2-AttB2; PNT3-AttB1 and PNT3-AttB2; PNT4-AttB1 and PNT4-AttB2.

The amplicons were then cloned into the pDest17 bacterial expression vector using the Gateway ® technology (Invitrogen). This vector allows expression of recombinant proteins under the control of the T7 promoter. The resulting protein is preceded by a stretch of 22 vector-encoded residues (MSYYHHHHHHLESTSLYKKAGS) encompassing a hexahistidine tag.

For the prokaryotic expression of PNT3 fused the green fluorescent protein (GFP), a 6His-tagged PNT3-GFP construct was generated by PCR amplification using pDest17/PNT3 as template and primers B1HisNT3 and NT3B2. After DpnI treatment, 1 μl of the first PCR was used as a template in a second PCR amplification using primers attl1a and attl2a as described in [90]. The second PCR product was then used in LR reaction with the expression vector pTH31 [91] to yield a construct referred to as His-PNT3-pTH31.

For the eukaryotic expression of PNT3-GFP, another 6His-tagged PNT3-GFP construct was generated by PCR using a GFP coding sequence optimized for eukaryotic expression. In a first PCR, the coding sequence of PNT3 was amplified using His-PNT3-pTH31 as template and primers HindNT3 and NT3R. In a second PCR, the coding sequence of GFP was PCR amplified using GFPq in pTTo [92] as template and primers GFPqF and GFPqXhoI. After DpnI treatment, 1 μl of PCR1 and 1 μl of PCR2 were used as overlapping megaprimers along with primers HindNT3 and GFPqXhoI in a third PCR. After purification, the third PCR product was digested by HindIII and XhoI and ligated into pCDNA3.1+ that had been digested with the same enzymes. The resulting construct is referred to as His-PNT3-GFPq-pCDNA3.1+.

For the eukaryotic expression of PNT3, a 6His-PNT3 construct was generated by PCR using His-PNT3-GFPq-pCDNA3.1+ as template and primers HindNT3 and NT3XhoI. After DpnI treatment, the PCR product was digested by HindIII and XhoI and ligated to pCDNA3.1+ as described above. The construct is referred to as His-PNT3-pCDNA3.1+.

For the prokaryotic expression of the PNT3 Y210A-Y211A-Y212A triple variant (PNT3^{3A}), the His-PNT3-pTH31 construct was used as template in two separate PCR amplifications using either primers attB1 and R_3ala-pNT3 (PCR1), or primers F_3ala-pNT3 and attB2 (PCR2). After DpnI treatment, 1 μl of PCR1 and 1 μl of PCR2 were used as overlapping megaprimers along with primers attB1 and attB2 in a third PCR. After purification, the third PCR product was inserted in pDONR by BP reaction (Invitrogen). After sequencing, the mutant coding sequence was transferred from pDONR to pDEST17 by LR reaction (Invitrogen). The resulting construct is referred to as mutPNT3-pDEST17.

For the eukaryotic expression of the PNT3 Y210A-Y211A-Y212A triple variant, the His-PNT3-pCDNA3.1+ construct was used as template in two separate PCR amplifications using either primers HindNT3 and R_3ala-pNT3 (PCR1), or primers F_3ala-pNT3 and NT3XhoI (PCR2). After DpnI treatment, 1 μl of PCR1 and 1 μl of PCR2 were used as overlapping megaprimers along with primers HindNT3 and NT3XhoI in a third PCR. After purification, the third PCR product was processed as described for generating His-PNT3-3A-pCDNA3.1+.

The constructs for the eukaryotic expression of GFP and 2OST (an irrelevant, control hexahistidine tagged protein) contain the ORF of GFPq [92] and of 2OST, respectively, cloned into pCDNA3.1+. The construct used for the prokaryotic expression of His-hsp70 has already been described [93].

Primers were all purchased from Eurofins Genomics. The sequence of all the primers is provided in **Supplementary Table S2**. All the constructs were verified by DNA sequencing (Eurofins Genomics) and found to conform to expectations.

Proteins expression and purification

The *E. coli* strain T7pRos was used for the expression of all the recombinant proteins. Cultures were grown over-night to saturation in LB medium containing 100 μg mL^{-1} ampicillin and 34 μg mL^{-1} chloramphenicol. An aliquot of the overnight culture was diluted 1/20 into 1 L of TB medium and grown at 37 °C 200 rpm. When the optical density at 600 nm (OD$_{600}$) reached 0.5-0.8, isopropyl β-D-thiogalactopyanoside (IPTG) was added to a final concentration of 0.5 mM, and the cells were grown at 25 °C 200 rpm overnight. The induced cells were harvested, washed and collected by centrifugation (5,000 rpm, 10 min). The resulting pellets were frozen at -20 °C.

Expression of ^{15}N-labelled PNT3 was performed as described above except that when the cultures reached an OD$_{600}$ of 0.6, the culture was centrifuged at 4,000 rpm for 10 min, and the pellet was resuspended in 250 mL of M9 medium (6 g L^{-1} of Na$_2$HPO$_4$, 3 g L^{-1} of KH$_2$PO$_4$, 0.5 g L^{-1} of NaCl, 0.246 g L^{-1} of MgSO$_4$) supplemented with 1 g L^{-1} of ^{15}NH$_4$Cl and 4 g L^{-1} of glucose. After one hour at 37 °C, IPTG was added to a final concentration of 0.5 mM, and the cells were subsequently grown at 25 °C overnight.

The purification protocol of the *Henipavirus* V proteins and of the HeV V ZnFD have been already reported [23], as was that of HeV PNT [8]. The PNT1-PNT4, PNT3^{3A} and PNT3-GFP proteins were purified as follows. The cellular pellet was resuspended (30 mL *per* liter of culture) in buffer A (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 1 mM phenyl methyl sulfonyl fluoride (PMSF), 0.1 mg mL^{-1} lysozyme, 10 μg mL^{-1} DNAse I and 20 mM MgSO$_4$. After an incubation of 20 min at 4 °C, the cells were disrupted by sonication (using a 750 W sonicator and 3 cycles of 30 s each at 45 % power output). Except for the PNT3-GFP protein, 6 M GuHCl was added to the lysate and the solution was incubated 1 hour at 4 °C with gentle agitation. The lysates were clarified by centrifugation at 14,000 g for 30 min at 4 °C and the supernatant was loaded onto a 5 mL Nickel column (GE Healthcare) pre-equilibrated in buffer A. The affinity resin was washed with 20 column volumes (CV) of buffer A. Proteins were eluted with buffer A supplemented with 250 mM imidazole. PNT3 and PNT3^{3A} were concentrated in the presence of 6 M GuHCl up to 1-2 mM using Centricon concentrators (Merk-Millipore) with a 10 kDa molecular mass cut-off, and then frozen at -20 °C.

Prior to each experiment, the proteins were subjected to size exclusion chromatography (SEC) that also enabled confirming homogeneity and monomeric nature of the proteins. The SEC column was equilibrated with buffer B (20 mM Tris/HCl pH 7.5, 100 mM NaCl, 5 mM EDTA). For samples to be used in CD, SAXS, EM and NMR experiments, SEC was carried out using a 50 mM phosphate pH 6.5, 5 mM EDTA buffer.

Fractions from all purification steps were analyzed by SDS-PAGE. Protein concentrations were estimated using the theoretical absorption coefficients at 280 nm as obtained using the program ProtParam from the EXPASY server.

The purity of the final purified products was assessed by SDS-PAGE (**Figure 1D**). The identity of the purified PNT3 and PNT3^{3A} samples was confirmed by mass spectrometry (MALDI-TOF) analysis of tryptic fragments obtained after digestion of the purified protein bands excised from SDS-polyacrylamide gels (**Supplementary Figure S4**). The bands were

then digested with trypsin (20 µl of trypsin/band at a concentration of 12.5 ng/µl). For each protein band, mass analyses were performed on a MALDI-TOF-TOF Bruker Ultraflex III spectrometer (Bruker Daltonics, Wissembourg, France) operating in reflector mode. The *m/z* range was from 600 to 4500, and α -Cyano-4-hydroxycinnamic acid was used as matrix solution. The mass standards were either autolytic tryptic peptides used as internal standards or peptide standards (Bruker Daltonics). Following MS analysis, MS/MS analyses were performed on the most intense peaks to identify the amino acid sequence of the protein band.

Hsp70 was purified according to [94] except that the last SEC step was replaced by a desalting step using a HiPrep 26/10 desalting column (GE Healthcare) and buffer B. The protein was then concentrated to 0.5 mM using Centricon concentrators (Merk-Millipore) with a 30 kDa molecular mass cut-off.

Turbidity and aggregation measurements

Turbidity was measured by monitoring the OD at 600 nm using a NanoDrop ND-1000 (Thermo Scientific) spectrophotometer. Samples (100 µL each), in the concentration range of 0-250 µM, were incubated 1-hour at RT either in the absence or in the presence of increasig concentrations of the crowding agent PEG_{300}. Different PEG (400, 4000, 6000, 8000) at different percentages (3-30 %) were also tested and the evolution of the droplets (400 nL) were monitored using a visible/UV light RockImager (Formulatrix). The formation of aggregates was monitored by measuring the OD at 340 nm of PNT3 samples incubated 1 hour either at RT or at 370C in the presence of increasing concentrations of salt.

Thioflavin T and Congo red binding assays

Thioflavin T (ThT) (Sigma-Aldrich) binding assays were performed using the dye at a final concentration of 20 µM in the presence of sodium azide 0.05 %. The samples were excited at 440 nm and emission was recorded between 450 and 500 nm. Fluorescence intensity was measured over time at each 15 min at 37 °C using a Cary Eclipse Fluorescence Spectrophotometer (Agilent). Data was plotted using Prism software and were analyzed using the Amylofit server (www.amylofit.ch.cam.ac.uk) [95].

CR binding assays were performed using the dye at a final concentration of 5 µM in the presence of 25 µM PNT3 (in sodium phosphate buffer at pH 6.5). The spectrum was recorded after 27 hours of incubation at 37°C using a NanoDrop ND-1000 (Thermo Scientific) spectrophotometer in the 400–700 nm range. Five µM CR in sodium phosphate buffer without PNT3 was used as a control.

Far-UV circular dichroism

CD spectra were measured using a Jasco 819 dichrograph, flushed with N_2 and equipped with a Peltier thermoregulation system. Proteins were loaded into a 1 mm quartz cuvette at 0.1 mg/mL (in 10 mM phosphate buffer at pH 6.5) and spectra were recorded at 20 °C. The scanning speed was 20 nm min^{-1}, with data pitch of 0.2 nm. Each spectrum is the average of three acquisitions. The spectrum of buffer was subtracted from the protein spectrum. Spectra were smoothed using the ''means-movement'' smoothing procedure implemented in the Spectra Manager package.

Mean molar ellipticity values per residue (MRE) were calculated as

$$[\theta] = \frac{3300 m \Delta A}{lcn} \qquad \text{(Eq. 1)}$$

where *l* is the path length in cm, *n* is the number of residues (133), *m* is the molecular mass in Daltons (15,138) and *c* is the concentration of the protein in mg mL^{-1}.

Small-angle X-ray scattering (SAXS)

SAXS measurements were carried out at the European Synchrotron Radiation Facility (Grenoble, France) on beamline BM29 (bending magnet) at a working energy of 12.5 KeV. Data were collected on a Pilatus (1M) detector. The wavelength was 0.992 Å. The sample-to-detector distance was 2.847 m, leading to scattering vectors (q) ranging from 0.028 to 4.525 nm^{-1}. The scattering vector is defined as $q = 4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The exposure time was optimized to reduce radiation damage. SAXS data were collected at 20 °C using purified protein samples (50 μL each). PNT3 concentrations were 1.0 and 2.0 mg mL^{-1}. Samples were incubated at 3°C and the data were collected at various times points from 0 to 630 min. Data were analyzed using the ATSAS program package [96]. Data reductions were performed using the established procedure available at BM29, and buffer background runs were subtracted from sample runs. The forward scattering intensities were calibrated using water as reference.

The R_g, forward intensity at zero angle I(0) and pair distance distribution function, P(r), were calculated with the program GNOM [97]. This enabled identifying the scattering curves corresponding to a monomeric species. For those latter curves, the R_g and forward intensity at zero angle I(0) were recalculated and refined with the program PRIMUS [98] according to the Guinier approximation at low q values, in a q.Rg range up to 1.3 :

$$Ln[I(q)] = Ln[I_0] - \frac{q^2 R_g^2}{3} \qquad \text{(Eq. 2)}$$

For the monomeric species at the highest protein concentration, we also attempted at describing it as conformational ensemble. To this end we used the program suite EOM 2.0 [42] using the default parameters.

The theoretical value of Rg (in Å) expected for an IDP was calculated using Flory's equation according to [41]:

Rg=$R_0 N^v$ \qquad (Eq. 3)

where N is the number of amino acid residues, R_0 is 2.54 ± 0.01 and v is 0.522 ± 0.01

Negative-staining transmission electron microscopy (TEM)

Drops of 2 μL of freshly purified PNT3 or PNT[3A] proteins (100-200 μM) either in the absence or in the presence of a two-fold molar excess of Hsp70, were deposited onto a glow-discharge carbon coated grid (Formwar/Carbon 300 mesh Cu, Agar Scientific). Prior to protein deposition, grids were exposed to plasma glow discharge for 20 seconds using a PELCO, easiGlow Glow Discharge Cleaning System (Ted Pella Inc. USA) (Current 25mA) in order to increase protein adhesion. To assess the fibrils stability, PNT3 (200 μM) supplemented with 5 mM SDS was deposited on the copper grids. The grids were washed three times with 20 μL of deionized water before incubating them for 1 min in a 2 % (w/v) Uranyl Acetate solution (LauryLab-France). Images were collected using a TECNAI T12 Spirit microscope (FEI company) operated at 120 kV and an Eagle 2Kx2K CCD camera (FEI company).

Nuclear Magnetic Resonance (NMR)

A sample of a freshly purified ^{15}N-labelled PNT3 at 0.1 mM in 50 mM phosphate pH 6.5, 5 mM EDTA, also containing D2O (10 %), was used for the acquisition of 1D ^{1}H and 2D ^{1}H,^{15}N HSQC spectra with an 22.3 T Bruker AvanceIII 950 ultra-shielded-plus spectrometer equipped with a triple resonance CryoProbe (TCI) at 310 K. The sample was incubated at 37°C and spectra were recorded at various time points. Spectra were recorded both at 288 and 310 K and were found to be superimposable.

X-ray diffraction

A sample of PNT3 at 20 μM from a ThT assay was centrifuged at 13000 rpm for 15 min and lyophilized. The powder was transferred into a mounted cryo-loop (Hampton Research). The fiber diffraction images were collected using a Microstar X-ray generator equipped with a MAR345 detector using a fixed wavelength of 1.54 Å and a sample-detector distance of 200 mm, with an exposure time of 300 s.

Assays of SDS Sensitivity

Preformed fibers of PNT3 (100 μM after 56 hours of incubation at 37°C) either non-supplemented or supplemented with 2% SDS, were passed through a 0.2 mm spin filter to remove fiber particles. UV absorbance, after subtraction of buffer contributions, was measured to monitor the amount of monomeric protein that passed through the filter.

Transfection of mammalian cells and viability assays

HEK293T cells were cultured in DMEM supplemented with 10% FBS at 37°C in a CO_2 incubator. Transfections were performed with PEI reagent (Polysciences) with a DNA:PEI ratio of 1:2 (w/w). CR staining and viability assay experiments were performed 72 and 48 hours after transfection, respectively. For CR staining, cells were fixed with 4 % PFA and incubated over night with 5 μM CR (Sigma) in a buffer containing 100 mM NaCl, 10 mM Hepes pH 7.4. Cells were then washed 1 hour with PBS diluted in water (1/4) and stored at 4°C in PBS before microscopy analysis. For cell viability assays, non-transfected or transfected (6His-PNT3, 6His-PNT3[3A], 6His-PNT3-GFP, GFP and 6His-2OST) cells were incubated 6 hours with H_2O_2 at different concentrations in triplicate in 96-well plates. PrestoBlue reagent (ThermoFisher) was added (10% v/v) to the 96-well plate and then incubated 2 additional hours at 37°C. Fluorescence was measured (ex 560 nm / em 590 nm) with a Saphire2 fluorimeter (Tecan). Cells expressing GFP or 6His-PNT3-GFP were observed under an epi-fluorescent microscope (Eclipse TS-100 Nikon).

Western Blot analysis

Expression 6His-PNT3, 6His-PNT33A, and 6His-2OST in HEK293T cells was assessed 48 hours post-transfection in 96-well plates. Cells were scrapped off after trypsinization (96 wells) and 5X Laemli loading buffer was added. Samples were boiled and 20 μ L of lysates were loaded onto a 15% SDS-PAGE gel. After electrophoresis, the gel was transferred on a nitrocellulose membrane and the proteins were stained using a Penta-His HRP conjugate (1/2000) (Biolegend). Signal was detected by chemiluminescence (Amersham ECL Western Blotting detecting reagent, GE Healthcare) using a Kodak, Image Station 440 luminometer.

References

[1] Wang LF, Yu M, Hansson E, Pritchard LI, Shiell B, Michalski WP, et al. The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. J Virol. 2000;74:9972-9.

[2] Gurley ES, Montgomery JM, Hossain MJ, Bell M, Azad AK, Islam MR, et al. Person-to-person transmission of Nipah virus in a Bangladeshi community. Emerg Infect Dis. 2007;13:1031-7.

[3] Homaira N, Rahman M, Hossain MJ, Epstein JH, Sultana R, Khan MS, et al. Nipah virus outbreak with person-to-person transmission in a district of Bangladesh, 2007. Epidemiol Infect. 2010;138:1630-6.

[4] Ching PK, de los Reyes VC, Sucaldito MN, Tayag E, Columna-Vingno AB, Malbas FF, Jr., et al. Outbreak of henipavirus infection, Philippines, 2014. Emerg Infect Dis. 2015;21:328-31.

[5] Hayman DT, Suu-Ire R, Breed AC, McEachern JA, Wang L, Wood JL, et al. Evidence of henipavirus infection in West African fruit bats. PLoS One. 2008;3:e2739.

[6] Bloyet LM, Welsch J, Enchery F, Mathieu C, de Breyne S, Horvat B, et al. HSP90 Chaperoning in Addition to Phosphoprotein Required for Folding but Not for Supporting Enzymatic Activities of Measles and Nipah Virus L Polymerases. J Virol. 2016;90:6642-56.

[7] Sourimant J, Rameix-Welti MA, Gaillard AL, Chevret D, Galloux M, Gault E, et al. Fine mapping and characterization of the L-polymerase-binding domain of the respiratory syncytial virus phosphoprotein. J Virol. 2015;89:4421-33.

[8] Habchi J, Mamelli L, Darbon H, Longhi S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. PLoS ONE. 2010;5:e11684.

[9] Blocquel D, Beltrandi M, Erales J, Barbier P, Longhi S. Biochemical and structural studies of the oligomerization domain of the Nipah virus phosphoprotein: Evidence for an elongated coiled-coil homotrimer. Virology. 2013;446:162-72.

[10] Beltrandi M, Blocquel D, Erales J, Barbier P, Cavalli A, Longhi S. Insights into the coiled-coil organization of the Hendra virus phosphoprotein from combined biochemical and SAXS studies. Virology. 2015;477:42-55.

[11] Bruhn-Johannsen JF, Barnett K, Bibby J, Thomas J, Keegan R, Rigden D, et al. Crystal structure of the Nipah virus phosphoprotein tetramerization domain. J Virol. 2014;88:758-62.

[12] Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, Tarbouriech N, et al. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. PLoS Pathog. 2013;9:e1003631.

[13] Habchi J, Blangy S, Mamelli L, Ringkjobing Jensen M, Blackledge M, Darbon H, et al. Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. J Biol Chem. 2011;286:13583-602.

[14] Habchi J, Martinho M, Gruet A, Guigliarelli B, Longhi S, Belle V. Monitoring structural transitions in IDPs by site-directed spin labeling EPR spectroscopy. Methods Mol Biol. 2012;895:361-86.

[15] Blocquel D, Habchi J, Gruet A, Blangy S, Longhi S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. Mol Biosyst. 2012;8:392-410.

[16] Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol. 2003;84:3239-52.

[17] Fontana JM, Bankamp B, Rota PA. Inhibition of interferon induction and signaling by paramyxoviruses. Immunol Rev. 2008;225:46-67.

[18] Audsley MD, Moseley GW. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. World J Virol. 2013;2:57-70.

[19] Park MS, Shaw ML, Munoz-Jordan J, Cros JF, Nakaya T, Bouvier N, et al. Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. J Virol. 2003;77:1501-11.

[20] Shaw ML, Garcia-Sastre A, Palese P, Basler CF. Nipah virus V and W proteins have a common STAT1-binding domain yet inhibit STAT1 activation from the cytoplasmic and nuclear compartments, respectively. J Virol. 2004;78:5633-41.

[21] Marsh GA, de Jong C, Barr JA, Tachedjian M, Smith C, Middleton D, et al. Cedar virus: a novel Henipavirus isolated from Australian bats. PLoS Pathog. 2012;8:e1002836.

[22] Ulane CM, Horvath CM. Paramyxoviruses SV5 and HPIV2 assemble STAT protein ubiquitin ligase complexes from cellular components. Virology. 2002;304:160-6.

[23] Salladini E, Delauzun V, Longhi S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. Mol Biosyst. 2017;13:2254-67.

[24] Uversky VN. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. Curr Opin Struct Biol. 2017;44:18-30.

[25] Holehouse AS, Pappu RV. Functional Implications of Intracellular Phase Transitions. Biochemistry. 2018;57:2415-23.

[26] Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. 2018.

[27] Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell. 2017;171:615-27 e16.

[28] Kato M, McKnight SL. A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. Annu Rev Biochem. 2017.

[29] Boeynaems S, Bogaert E, Kovacs D, Konijnenberg A, Timmerman E, Volkov A, et al. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell. 2017;65:1044-55 e5.

[30] Maharana S, Wang J, Papadopoulos DK, Richter D, Pozniakovsky A, Poser I, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science. 2018.

[31] Lin Y, Currie SL, Rosen MK. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J Biol Chem. 2017;292:19110-20.

[32] Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. Elife. 2018;7.

[33] Li HR, Chiang WC, Chou PC, Wang WJ, Huang JR. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. J Biol Chem. 2018;293:6090-8.

[34] Galzitskaya OV. Repeats are one of the main characteristics of RNA-binding proteins with prion-like domains. Mol Biosyst. 2015;11:2210-8.

[35] Axelrod D, Koppel DE, Schlessinger J, Elson E, Webb WW. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J. 1976;16:1055-69.

[36] Klunk WE, Pettegrew JW, Abraham DJ. Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. J Histochem Cytochem. 1989;37:1273-81.

[37] LeVine H, 3rd. Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. Protein Sci. 1993;2:404-10.

[38] Sunde M, Blake C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Protein Chem. 1997;50:123-59.

[39] Stromer T, Serpell LC. Structure and morphology of the Alzheimer's amyloid fibril. Microsc Res Tech. 2005;67:210-7.

[40] Morris KL, Serpell LC. X-ray fibre diffraction studies of amyloid fibrils. Methods Mol Biol. 2012;849:121-35.

[41] Bernado P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J. 2009;97:2839-45.

[42] Tria G, Mertens HDT, Kachala M, Svergun D. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ. 2015;2:202-17.

[43] Dobson CM. Protein folding and misfolding. Nature. 2003;426:884-90.

[44] Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. Annu Rev Biochem. 2017;86:27-68.

[45] Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell. 2009;137:146-58.

[46] Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149:753-67.

[47] Fenyi A, Coens A, Bellande T, Melki R, Bousset L. Assessment of the efficacy of different procedures that remove and disassemble alpha-synuclein, tau and A-beta fibrils from laboratory material and surfaces. Sci Rep. 2018;8:10788.

[48] Klucken J, Shin Y, Masliah E, Hyman BT, McLean PJ. Hsp70 Reduces alpha-Synuclein Aggregation and Toxicity. J Biol Chem. 2004;279:25497-502.

[49] Dedmon MM, Christodoulou J, Wilson MR, Dobson CM. Heat shock protein 70 inhibits alpha-synuclein fibril formation via preferential binding to prefibrillar species. J Biol Chem. 2005;280:14733-40.

[50] Patterson KR, Ward SM, Combs B, Voss K, Kanaan NM, Morfini G, et al. Heat shock protein 70 prevents both tau aggregation and the inhibitory effects of preexisting tau aggregates on fast axonal transport. Biochemistry. 2011;50:10300-10.

[51] Aprile FA, Arosio P, Fusco G, Chen SW, Kumita JR, Dhulesia A, et al. Inhibition of alpha-Synuclein Fibril Elongation by Hsp70 Is Governed by a Kinetic Binding Competition between alpha-Synuclein Species. Biochemistry. 2017;56:1177-80.

[52] Xue YL, Wang H, Riedy M, Roberts BL, Sun Y, Song YB, et al. Molecular dynamics simulations of Hsp40 J-domain mutants identifies disruption of the critical HPD-motif as the key factor for impaired curing in vivo of the yeast prion [URE3]. J Biomol Struct Dyn. 2018;36:1764-75.

[53] Lahaye X, Vidy A, Pomier C, Obiang L, Harper F, Gaudin Y, et al. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. J Virol. 2009;83:7948-58.

[54] Heinrich BS, Cureton DK, Rahmeh AA, Whelan SP. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. PLoS Pathog. 2010;6:e1000958.

[55] Hoenen T, Shabman RS, Groseth A, Herwig A, Weber M, Schudt G, et al. Inclusion bodies are a site of ebolavirus replication. J Virol. 2012;86:11779-88.

[56] Zhang S, Chen L, Zhang G, Yan Q, Yang X, Ding B, et al. An amino acid of human parainfluenza virus type 3 nucleoprotein is critical for template function and cytoplasmic inclusion body formation. J Virol. 2013;87:12457-70.

[57] Rincheval V, Lelek M, Gault E, Bouillier C, Sitterlin D, Blouquit-Laye S, et al. Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. Nat Commun. 2017;8:563.

[58] Cifuentes-Munoz N, Branttie J, Slaughter KB, Dutch RE. Human Metapneumovirus Induces Formation of Inclusion Bodies for Efficient Genome Replication and Transcription. J Virol. 2017;91.

[59] Netherton CL, Wileman T. Virus factories, double membrane vesicles and viroplasm generated in animal cells. Curr Opin Virol. 2011;1:381-7.

[60] Nikolic J, Le Bars R, Lama Z, Scrima N, Lagaudriere-Gesbert C, Gaudin Y, et al. Negri bodies are viral factories with properties of liquid organelles. Nat Commun. 2017;8:58.

[61] Murakami T, Qamar S, Lin JQ, Schierle GS, Rees E, Miyashita A, et al. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron. 2015;88:678-90.

[62] Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell. 2015;162:1066-77.

[63] Elbaum-Garfinkle S, Brangwynne CP. Liquids, Fibers, and Gels: The Many Phases of Neurodegeneration. Dev Cell. 2015;35:531-2.

[64] Alonso LG, Garcia-Alai MM, Smal C, Centeno JM, Iacono R, Castano E, et al. The HPV16 E7 viral oncoprotein self-assembles into defined spherical oligomers. Biochemistry. 2004;43:3310-7.

[65] Smal C, Alonso LG, Wetzler DE, Heer A, de Prat Gay G. Ordered self-assembly mechanism of a spherical oncoprotein oligomer triggered by zinc removal and stabilized by an intrinsically disordered domain. PLoS One. 2012;7:e36457.

[66] Dantur K, Alonso L, Castano E, Morelli L, Centeno-Crowley JM, Vighi S, et al. Cytosolic accumulation of HPV16 E7 oligomers supports different transformation routes for the prototypic viral oncoprotein: the amyloid-cancer connection. Int J Cancer. 2009;125:1902-11.

[67] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science. 2017;357.

[68] Protter DS, Parker R. Principles and Properties of Stress Granules. Trends Cell Biol. 2016;26:668-79.

[69] Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell. 2016;164:487-98.

[70] Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc Natl Acad Sci U S A. 2011;108:4334-9.

[71] Oglesbee M, Kim MY, Shu Y, Longhi S. Extracellular HSP70, Neuroinflammation and Protection against Viral Virulence. In: Asea AAAaK, P, editor. Chaperokine activity of heat shock proteins: Springer; 2018.

[72] Kim MY, Shu Y, Carsillo T, Zhang J, Yu L, Peterson C, et al. hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain. J Virol. 2013;87:998-1009.

[73] Oglesbee MJ, LaBranche T. Inside-out: extracellular roles for heat shock proteins. Vet Pathol. 2013;50:921-4.

[74] Kim MY, Ma Y, Zhang Y, Li J, Shu Y, Oglesbee M. hsp70-dependent antiviral immunity against cytopathic neuronal infection by vesicular stomatitis virus. J Virol. 2013;87:10668-78.

[75] Ma Y, Duan Y, Wei Y, Liang X, Niewiesk S, Oglesbee M, et al. Heat shock protein 70 enhances mucosal immunity against human norovirus when coexpressed from a vesicular stomatitis virus vector. J Virol. 2014;88:5122-37.

[76] Bourhis JM, Canard B, Longhi S. Structural disorder within the replicative complex of measles virus: functional implications. Virology. 2006;344:94-110.

[77] Longhi S, Oglesbee M. Structural disorder within the measles virus nucleoprotein and phosphoprotein. Protein and Peptide Letters. 2010;17:961-78.

[78] Leyrat C, Gerard FC, de Almeida Ribeiro E, Jr., Ivanov I, Ruigrok RW, Jamin M. Structural disorder in proteins of the rhabdoviridae replication complex. Protein Pept Lett. 2010;17:979-87.

[79] Longhi S. Structural disorder within the measles virus nucleoprotein and phosphoprotein: functional implications for transcription and replication. In: Luo M, editor. Negative strand RNA virus. Singapore: World Scientific Publishing; 2011. p. 95-125.

[80] Ivanov I, Yabukarski F, Ruigrok RW, Jamin M. Structural insights into the rhabdovirus transcription/replication complex. Virus Res. 2011;162:126-37.

[81] Leyrat C, Schneider R, Ribeiro EA, Jr., Yabukarski F, Yao M, Gerard FC, et al. Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. J Mol Biol. 2012;423:182-97.

[82] Blocquel D, Bourhis JM, Eléouët JF, Gerlier D, Habchi J, Jamin M, et al. Transcription et réplication des Mononégavirales: une machine moléculaire originale. Virologie. 2012;16:225-57.

[83] Habchi J, Mamelli L, Longhi S. Structural disorder within the nucleoprotein and phosphoprotein from measles, Nipah and Hendra viruses. In: Uversky VN, Longhi S, editors. Flexible viruses: structural disorder in viral proteins. Hoboken, New Yersey: John Wiley and Sons; 2012. p. 47-94.

[84] Habchi J, Longhi S. Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. Mol Biosyst. 2012;8:69-81.

[85] Communie G, Ruigrok RW, Jensen MR, Blackledge M. Intrinsically disordered proteins implicated in paramyxoviral replication machinery. Curr Opin Virol. 2014;5:72-81.

[86] Leung DW, Borek D, Luthra P, Binning JM, Anantpadma M, Liu G, et al. An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. Cell Rep. 2015;11:376-89.

[87] Erales J, Blocquel D, Habchi J, Beltrandi M, Gruet A, Dosnon M, et al. Order and disorder in the replicative complex of paramyxoviruses. Adv Exp Med Biol. 2015;870:351-81.

[88] Habchi J, Longhi S. Structural Disorder within Paramyxoviral Nucleoproteins and Phosphoproteins in Their Free and Bound Forms: From Predictions to Experimental Assessment. Int J Mol Sci. 2015;16:15688-726.

[89] Longhi S, Bloyet LM, Gianni S, Gerlier D. How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. Cell Mol Life Sci. 2017;74:3091-118.

[90] Gruet A, Dosnon M, Blocquel D, Brunel J, Gerlier D, Das RK, et al. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. FEBS J. 2016;283:576-94.

[91] van den Berg S, Lofdahl PA, Hard T, Berglund H. Improved solubility of TEV protease by directed evolution. J Biotechnol. 2006;121:291-8.

[92] Durocher Y, Perret S, Kamen A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res. 2002;30:E9.

[93] Zhang X, Bourhis JM, Longhi S, Carsillo T, Buccellato M, Morin B, et al. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. Virology. 2005;337:162-74.

[94] Couturier M, Buccellato M, Costanzo S, Bourhis JM, Shu Y, Nicaise M, et al. High Affinity Binding between Hsp70 and the C-Terminal Domain of the Measles Virus Nucleoprotein Requires an Hsp40 Co-Chaperone. J Mol Recognit. 2010;23:301-15.

[95] Meisl G, Kirkegaard JB, Arosio P, Michaels TC, Vendruscolo M, Dobson CM, et al. Molecular mechanisms of protein aggregation from global fitting of kinetic models. Nat Protoc. 2016;11:252-72.

[96] Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Cryst 2012;45:342-50.

[97] Svergun D. Determination of the regularization parameters in indirect-trasform methods using perceptual criteria. J Appl Cryst. 1992;25:495-503.

[98] Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Cryst. 2003;36:1277-82.

3.4 Interaction study between IDPs

Understanding protein-protein interaction involving IDPs is a central theme for experimental and theoretical studies of IDPs. Particularly challenging are the cases of extreme fuzziness of the complex, which is now appearing as a common interaction mode, requiring new models for their description.[25,26,108] So far, NMR spectroscopy has proven to be one of the most powerful techniques to characterize at atomic level intrinsically disordered proteins and their interactions, including those cases where the formed complexes are highly dynamics. An interesting case is the one of hub proteins. Often, these proteins show elevated level of disorder, as the flexibility helps promiscuous interactions in one-to-many mode. Disorder appears to be a very useful feature for hub viral proteins that have to interact with multiple partners to promote viral replication. A valuable example is the E1A protein, which recently has been characterized at atomic resolution.[109] The E1A protein is the most extensively studied viral transcriptional regulator and can be considered a prototypic model.[30,47] Through its interaction with over 50 cellular factors E1A interfere with cell cycle regulation by binding key cellular proteins. In particular, association of E1A with CREB binding protein CBP, or its paralog p300, results in global hypoacetylation of K18 of histone H3 and repress differentiation. CBP is a transcriptional co-regulator able to interact with a large number of proteins through multiple protein-interaction domains and with acetyl-transferase activity, playing critical roles in basic cellular processes. This very large protein presents seven folded domains that have been structurally characterized in recent years[110–114] involved in several functional interaction with transcription factors. Although the interaction between E1A and CBP have been intensively investigated, only in the last decade structural characterization of E1A in the free state and fragments of E1A with folded domains of CBP have been characterized using NMR spectroscopy.[109,115,116] On the other hand, the regions outside the globular domains are predicted to be intrinsically disordered, constituting nearly 60% of the protein and thus also the disordered regions may have an active role in these numerous interaction far beyond the linker role.

The study of the interaction between E1A and the disordered region contained between the two folded domain involved in E1A-CBP interaction, called ID4[117], have different interesting aspect. First of all, interaction between disordered

proteins are still poorly investigated and thus their characterization is important to elucidate the binding driving forces. Despite the large abundance of disordered regions, the role of disorder in transcription factor like CBP is an open question. Furthermore, as many viral proteins are characterized by high level of disorder is important to understand the role of this feature in viral infections.

NMR spectroscopy allows us to investigate this flexible system providing information about dynamic variation that can be important to drive the mechanism of interaction. Many viral IDPs interfere with cell cycle regulation exploiting multiple weak interactions at once, creating high affinity for the partner. These weak interactions are very hard to discover using other techniques, while NMR allows to highlight every little change in the systems, providing information at atomic resolution about the residue directly involved in the interaction.[118–120]

Of course, understanding the mechanism of functional interaction responsible for viral replication and the responsible binding site could provide an excellent starting point to develop new strategy to prevent these viral infections.

3.4.1 Article: Fuzzy interaction between intrinsically disordered proteins: an NMR study

Fuzzy interaction between intrinsically disordered proteins: an NMR study

Maria Grazia Murrali, Isabella C. Felli, Roberta Pierattelli

Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy.

ASBTRACT

Direct interaction between intrinsically disordered proteins is often difficult to be characterized hampering the elucidation of their binding mechanism. Particularly challenging is the study of fuzzy complexes, in which the intrinsically disordered proteins or regions retain conformational freedom within the assembly. So far, nuclear magnetic resonance (NMR) spectroscopy has proven to be one of the most powerful techniques to characterize at atomic level intrinsically disordered proteins and their interactions, including those cases where the formed complexes are highly dynamic. Here we present the characterization of the interaction between a viral protein, the Early region 1A protein from Adenovirus (E1A) and a disordered region of the human CREB-binding protein, namely the fourth intrinsically disordered linker CBP-ID4. E1A has been widely studied as a prototypical viral oncogene. Its interaction with two folded domains of CBP has been mapped before, proving hints to understand some functional aspects of the interaction with this transcriptional coactivator. However, the role of the linkers of CBP in this interaction has never been investigated before.

INTRODUCTION

The Adenovirus (AdV) is responsible for respiratory and gastric infections in humans and oncogenic transformation in rodents. The early region 1A (E1A) gene is the first expressed following adenoviral infection and it plays a central function in viral replication, activating genes required for viral replication and for interfering with the host cell cycle regulation.[1,2] The E1A gene encodes two major proteins of 289 and 243 amino acids (E1A13S and E1A12S respectively) early after infection. Sequence comparisons of the E1A13S of several adenovirus serotypes display the presence of four conserved regions, termed CR1 to CR4, which are separated by less conserved regions.[3,4] There are five regions considered relevant for E1A13S function: the N-terminus (aa. 1-41); CR1 (aa. 42-72); CR2 (aa. 115-137); CR3 (aa. 144-191); CR4 (aa. 240-288). Interestingly, E1A13S and E1A12S are identical except for the CR3, a module of 46 amino acids, that belongs only to the largest of the two. E1A is considered a key player in viral replication through transcriptional regulation of other viral genes and contains two highly conserved CXXC motifs that are responsible for Zn^{II} binding.[5] The two E1A splicing variants have been characterized by NMR spectroscopy, showing that they are both highly disordered.[6]Only the CR3 region presents a minimal fold composed by two α-helices of less than 10 amino acids, embedded in an highly flexible polypeptide.[6]

Viruses are obligate intracellular parasites, completely dependent on host cell functions. By targeting multiple cellular hubs, viral proteins can reprogram virtually all mechanisms of cells biology, rendering the infected cell an amenable environment for virus replication.[7] The high flexibility of E1A is one of its crucial features, allowing it to adopt multiple conformations and to bind to diverse partners maximizing its interaction capabilities.[8] E1A participates in important protein-protein interactions with over 50 cellular factors, deregulating the cell cycle by interfering with key cellular proteins[2]. One of the most important, and most studied, interaction of E1A is the one with Cyclic-AMP-Response-Element-Binding protein (CREB)-Binding Protein (CBP) and its paralog p300.[9] The CBP/p300 is a transcriptional coactivator [10], able to interact with a large number of proteins through multiple protein-interaction domains and with acetyl-transferase activity, playing critical roles in basic cellular processes.[11] This 2442 residue long protein presents seven folded domains that have been structurally characterized in recent years[12–16] separated by five disordered linkers (denoted ID#, where # is the number of the linker).[17] Interestingly, the regions outside the globular domains are

predicted to be intrinsically disordered, constituting nearly 60% of the protein composition. Some of the five disordered linkers have been structurally characterized at atomic resolution, but little is known about their functional roles.[17,18]

The interaction between E1A and CBP has been reported in several studies[19,20] revealing that two folded domains of CBP,[2,11,21] the transcriptional adaptor zinc finger-2 domain (TAZ2) and the nuclear coactivator binding domains (NCBD), are the main responsible for binding. The TAZ2 domain binds specifically to the acidic transactivation domains of several transcription factors, including the p53 tumour suppressor,[19–21] ETF2 and members of the STAT and C/EBP families.[22–24] The interaction of E1A with this domain can inhibit the histone acetyltransferase activity and prevents interactions with other transcription factors. Association of E1A with CBP/p300 results in global hypoacetylation of K18 of H3 and may be linked to the ability of E1A to induce oncogenic transformations.[1] Moreover, E1A can compete with cellular factors for binding to NCBD, being able to repress differentiation and the immune response to viral infection.[16]

The interaction of fragments of E1A with the TAZ2 and NCBD single domains was also investigated by NMR spectroscopy. [25,26]These two globular domains of CBP involved in the interaction with E1A are connected by the disordered linker ID4. This 207 residue long linker is largely disordered but presents two partially populated helical fragments that are well conserved, suggesting their possible functional role.[27] Hence, we decided to investigate whether ID4 has an active role in the functional interaction between E1A and CBP, providing the first structural and dynamic characterization by NMR spectroscopy of the interaction between the entire E1A12S protein and a potential target.

MATERIAL AND METHODS

The ^{15}N-labeled and unlabeled E1A12S and ^{15}N labeled and unlabeled E1ACR3, were prepared as previously described.[6]. The NMR samples were prepared exchanging the final buffer in a degassed water buffer containing 20 mM TRIS, 50 mM KCl, 1 mM TCEP at pH=7.4, with 10% D_2O added for lock signal.

The $^{13}C,^{15}N$ labeled and unlabelled ID4 was prepared as previously described.[17] The NMR samples were prepared exchanging the final buffer in a degassed water buffer containing 20 mM TRIS, 50 mM KCl, 1 mM TCEP at pH=7.4, with 10% D_2O added for lock signal.

For all the titrations, the solutions of labelled proteins were split equally into two samples. One of these was mixed with the unlabeled titrant to obtain a 1:2 complex and the other was mixed with the same volume of NMR buffer. These two solutions were used to prepare the various samples with different molar ratios of the two proteins.

Five points per titration were recorded. The ^1H-^{15}N FHSQC[28] or ^1H-^{15}N BEST-TROSY[29] spectra were acquired at 283 K on a 22.3 T Bruker AVANCE III spectrometer, operating at 950 MHz ^1H, 238.8 MHz ^{13}C and 96.32 MHz ^{15}N equipped with a cryogenically cooled probehead in 3 mm NMR tubes.

The ^{13}C detected experiments were acquired at 16.4 T Bruker AVANCE NEO spectrometer, operating at 700.06 MHz ^1H, 176.05 MHz ^{13}C and 70.97 MHz ^{15}N equipped with a direct triple-resonance (TXO) cryogenically cooled probehead optimized for ^{13}C detection. Titration points were recorded in 5 mm shigemi NMR tube, acquiring 2D CON[30], 2D (HA)CONpro[31], 2D (HN)CON[32].

All the heteronuclear relaxation experiments (^{15}N R_1, ^{15}N R_2 and ^1H–^{15}N NOE)[33–35] were acquired using E1A12S at about 150 µM concentration and E1A12S + ID4 1:1 complex or ^{15}N labeled ID4 at about 200 µM concentration and ID4 + E1A12S 1:1 complex. The spectra were recorded at 283 K on a 16.4 T Bruker AVANCE NEO 700 spectrometer operating at 700.06 MHz ^1H, 176.05 MHz ^{13}C and 70.97 MHz ^{15}N equipped with an inverse triple resonance (TCI) cryogenically cooled probehead.

The ^{15}N R_1 and R_2 experiments were acquired with 8 scans per increment (2048 x 256 points) and a relaxation delay of 3.0 s. To determine the ^{15}N R_1 the following delays were used: 20 ms, 60 ms, 120 ms, 180 ms, 240 ms, 400 ms, 500 ms, 600 ms, 800 ms, 1 s, 1.2 s. To determine the ^{15}N R_2 the following delays were used: 32 ms, 64 ms, 96 ms , 128 ms, 160 ms, 190 ms, 260 ms, 320 ms, 380 ms, 440 ms, 500 ms. The ^1H–^{15}N NOEs experiments were acquired with 48 scans (2048 x 256 points) and a relaxation delay of 6.0 s.

The ^{15}N relaxation rates (R_1 and R_2) were determined by fitting the cross-peak intensities measured as a function of the variable delay, to single-exponential decay curves using the

Bruker Dynamic Center 2.4, available as stand-alone ancillary software of TopSpin by Bruker. ^{1}H-^{15}N NOE values were obtained as the ratio between peak intensities in spectra recorded with and without ^{1}H saturation. All the spectra were acquired, processed, and analysed by using Bruker TopSpin 3.5pl7 or Bruker TopSpin 4.0.1 software.

CD spectra to assess the stability of the protein secondary structure were acquired in the far-UV with a Jasco J-810 spectropolarimeter using a 0.1 cm path length quartz cuvette. The mean of 5 scans between 190 and 250 nm wavelength was calculated by subtraction of the corresponding buffer spectrum.

RESULTS and DISCUSSION

The human Adv E1A12S protein is an intrinsically disordered protein of 243 amino acids. E1A is the most extensively studied viral transcriptional regulator but, despite the large number of studies about E1A interactions, the structural characterization at atomic level of the entire protein was obtained only in 2016 exploiting NMR spectroscopy.[6] The NMR characterization unveils the high disorder of E1A12S, but with a heterogeneous behaviour in terms of structural and dynamic properties. This protein presents a significant α-helical secondary structure propensity encompassing residues 12-22 in the N-terminal part; a large number of proline residues (20% of the total residues) are uniformly distributed along the polypeptide.

Residues 1-25 in the N-terminus and a motif between residues 54 and 83 that contains a fragment of CR1 and 10 hydrophobic residues well conserved among E1A serotypes[25] were identified as interaction sites with TAZ2.[21] The NMR studies carried out on two short fragments of E1A (aa. 1-36 and aa. 53-91), reported the folding upon binding of these regions, with the formation of α-helical secondary structure in the CR1 region.[25]

CBP-ID4 is a 207 residue long protein (CBP residues 1851-2057) located between the TAZ2 and NCBD domains of CBP. This disordered linker is rich in positively charged amino acids and presents a large number of proline residues (22% of the total residues). Proline residues are uniformly distributed along the sequence, except for two regions that reveal two partially populated helical fragments termed Helix I (aa. 2-25, aa. 1852-1875 of CBP) and Helix II (aa. 101-128, aa. 1951-1978)[17]. The potential functional importance of these two regions is

suggested by the higher level of conservation with respect to the more disordered part[36] and we hypothesized a possible role of these helical regions in the interaction with E1A.

NMR spectroscopy has the potential of probing the more dynamic aspects of IDPs interactions, allowing the characterization at atomic resolution of highly dynamic and low affinity complexes. This is particularly important to understand the specificity of IDPs towards their partner proteins, the potential inter-play between different linear motifs and, in general, the regulatory mechanisms associated with multivalent interactions.

Mapping the interaction between E1A12S and CBP-ID4

The first step was the titration of ^{15}N labelled E1A12S with unlabelled CBP-ID4 (Figure 1). Upon addition of ID4 there are several changes in the ^{1}H-^{15}N correlation spectra. In particular, the first half of the sequence of E1A is perturbed by the interaction, showing different binding sites with different chemical exchange rates: a large group of peaks disappeared due to an exchange regime that causes broadening of the signals, while other peaks undergo variation in nuclear chemical shifts, as result of a fast exchange regime between free and bound state. Despite a general decrease of peak intensities affecting all the signals of the protein, the two effects described above are more pronounced for the first hundred residues with a prominent effect in two regions: the N–terminus (aa.2-32) and in the CR1 (aa. 66-81) (Figure 2A). The extension of the effect is confirmed also by the chemical shift perturbation observed for the residues in fast exchange, shown in the Garrett plot[37,38] for the 1:1 complex (Figure 2B).

On the 1:1 adduct, ^{15}N R_1 and R_2 relaxation rates and the ^{1}H-^{15}N NOE were measured (Figure 3) and compared with those of the free protein. The R_2 of the complex were significantly higher than those of free E1A12S in the first part of the protein, while the ^{1}H-^{15}N NOE and the R_1 were only slightly varied. In the plot of R_2 the missing data derive from residues that are not observable, not resolved of belonging to proline residues.

Mapping the interaction between CBP-ID4 and E1A12S

The titration of ^{15}N CBP-ID4 with unlabelled E1A12S showed no chemical shift perturbations of the CBP-ID4 signals in ^{1}H-^{15}N correlation spectra, but several signals disappeared or lost intensity progressively by increasing the concentration of E1A12S (Figure 4). Most of the peaks not present in the spectrum of the adduct belong to the first part of the protein, suggesting a

possible involvement of the first α-helical region in the interaction. The extent of overlap in the spectra, however, prevented a less coarse description of the binding site.

To refine the extension of the binding site, we exploited ^{13}C detection experiments,[39] which provide increased dispersion of nuclear chemical shifts and enable detection of proline residues in a straightforward manner.[31] The simplest ^{13}C-detected experiment is the 2D CON that allows to detect the signals that correlate the backbone carbonyl of one amino acid (C'$_i$) with the amidic nitrogen of the neighbouring peptide bond (N$_{i+1}$).[40] To acquire the C'$_i$-N$_{i+1}$ correlation several experimental variants can be used, exploiting C', H$^\alpha$ or H^N as starting magnetization source, which present different features in terms of sensitivity, time of acquisition and detected signals.[30,32,41] In the C'-start 2D CON of the 1:1 adduct, a large number of peaks belonging to residues in the two α-helical regions, are not present. However, exploiting the H^N CON, which is more sensitive, it appeared that most of the missing peaks belong to residues in Helix I, while the majority of peaks of residues in Helix II remain unchanged. (Figure 5). Exploiting ^{13}C-detection it was possible also to highlight the role of the proline residues in this interaction.[31] Out of the many proline residues present in CBP-ID4, only Pro 28, located at the end of Helix I, is affected by the interaction, confirming Helix I as main target of E1A.

To investigate the dynamic properties of the 1:1 complex of CBP-ID4 and E1A12S, the [15]N R$_1$, [15]N R$_2$ and [1]H-^{15}N NOE were measured (Figure 6). The missing values in the [15]N R$_2$ plot are due to broadening of the signals in the spectrum, while the R$_2$ of the proline residues were measured using ^{13}C-detected experiments.

In summary, the effect of the interaction on the dynamics of the two proteins points to a slight increase in [15]N R$_2$ values spread all over the two polypeptides, with higher increase in the regions showing α-helical propensity. This effect can be related to the decreased flexibility of the two systems in the complex with respect to the two isolated forms. Nevertheless, there are no indications of the occurrence of significant changes in conformation in any of the two proteins. This was confirmed by circular dichroism analysis of the two separated proteins and of the complex (not shown), suggesting that the complex maintains an overall general flexibility typical of fuzzy complexes.

Mapping the interaction between E1ACR3 and CBP-ID4

The interaction between CBP and E1ACR3 has not been mapped at atomic level, although direct binding of CBP to CR3 has been detected with purified proteins and by co-precipitation experiment of CBP and E1A13S cleaved of the N-terminus.[42] Both the E1A N-terminus/CR1 and CR3 function as independent transactivation domains, but interestingly the relative abundance of E1A12S and E1A13S is able to tune the overall transcription processes. As the CR3 is the only difference between E1A12S and E1A13S, the effect of tuning of E1A12S is related to its capability to interfere with CR3-transcription activation. It is reported to interact with a wide variety of different transcription factors, including CBP/p300.[42–45] CBP/p300 is the only factor sequestered by N-terminus of E1A12S but it is sufficient to deplete the CR3 transcription activation. Thus, it is important to investigate the competition between N-terminus/CR1 of E1A12S and CR3 in E1A13S, to understand the mechanism of the transcriptional activation of E1A. We thus decided to study the interaction between E1ACR3 and CBP-ID4.

E1ACR3 is a conserved region present in E1A13S starting at residue 144. This fragment of 46 amino acids presents two α-helices beginning with CXXC motifs that comprise residues 153-165 and 170-180 (E1A13S numbering). It present five cysteine residues, four of which are involved in Zn^{II} binding. E1ACR3 appears to be crucial for E1A function, as alteration of CR3 generally abolishes E1A transactivation.[42]

The NMR analysis of the titration of [15]N E1ACR3 with CBPID4 showed some signals chemical shifts perturbations and intensities variations (Figure 7). The complementary titration (^{15}N CBP-ID4 with E1ACR3) displayed that the binding site on CBP-ID4 is located in Helix I, with broadening of CBP-ID4 signals upon increasing the concentration of E1ACR3 (Figure 8). Thus, these two proteins compete for the same binding site on the disordered region of CBP and this may have importance in transactivation regulation.

CONCLUSIONS

In this study, the interaction of the entire E1A12S protein with the ID4 fragment of CBP has been mapped at atomic resolution. The importance of E1A12S in viral infection and in the modulation of viral transcription renders this disordered protein a very important and

complex system to be investigated. The use of NMR spectroscopy to investigate disordered proteins has proved to be once again fundamental to discover the formation of very flexible complexes, being able to detect multiple weak local interactions in order to form a stable high affinity interaction.

It was previously reported that E1A uses multiple molecular recognition features (MoRFs)[46,47] that confer high specificity but low affinity to the interaction with target proteins, resulting in a stable high affinity interaction. For example, the FXDXXXL motif (aa. 66-72) of E1A is able to interact with the TAZ2 and is also found in other viral and cellular proteins that bind CBP. This fragment is also involved in the interaction with CBP-ID4, and the latter might have an active role in mediating the interaction with TAZ2. Further studies are on-going to understand the role of E1ACR3 in targeting CBP.

REFERENCES

1. Frisch, S. M. & Mymryk, J. S. Adenovirus-5 E1A: paradox and paradigm. *Nat. Rev. Mol. Cell Biol.* **3**, 441–452 (2002).

2. Pelka, P., Ablack, J. N. G., Fonseca, G. J., Yousef, A. F. & Mymryk, J. S. Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes. *Journal of Virology* **82**, 7252–7263 (2008).

3. Hearing, P. & Shenk, T. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. *Cell* **33**, 695–703 (1983).

4. Avvakumov, N., Kajon, A. E., Hoeben, R. C. & Mymryk, J. S. Comprehensive sequence analysis of the E1A proteins of human and simian adenoviruses. *Virology* **329**, 477–492 (2004).

5. Culp, J. S. *et al.* The 289-amino acid E1A protein of adenovirus binds zinc in a region that is important for trans-activation. *Proc Natl Acad Sci U S A* **85**, 6450–6454 (1988).

6. Hošek, T. *et al.* Structural and Dynamic Characterization of the Molecular Hub Early Region 1A (E1A) from Human Adenovirus. *Chemistry* **22**, 13010–13013 (2016).

7. Dyson, H. J. & Wright, P. E. How Do Intrinsically Disordered Viral Proteins Hijack the Cell? *Biochemistry* **57**, 4045–4046 (2018).

8. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends in Biochemical Sciences* **36**, 159–169 (2011).

9. O'Connor, M. J., Zimmermann, H., Nielsen, S., Bernard, H. U. & Kouzarides, T. Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. *J. Virol.* **73**, 3574–3581 (1999).

10. Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).

11. Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J. Biol. Chem.* **291**, 6714–6722 (2016).

12. De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. *J. Mol. Biol.* **303**, 243–253 (2000).

13. De Guzman, R. N., Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* **44**, 490–497 (2005).

14. Brüschweiler, S., Konrat, R. & Tollinger, M. Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem. Biol.* **8**, 1600–1610 (2013).

15. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521–534 (2004).

16. Kjaergaard, M., Andersen, L., Nielsen, L. D. & Teilum, K. A folded excited state of ligand-free nuclear coactivator binding domain (NCBD) underlies plasticity in ligand recognition. *Biochemistry* **52**, 1686–1693 (2013).

17. Piai, A. *et al.* Just a Flexible Linker? The Structural and Dynamic Properties of CBP-ID4 Revealed by NMR Spectroscopy. *Biophys. J.* **110**, 372–381 (2016).

18. Contreras-Martos, S. *et al.* Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci Rep* **7**, 4676 (2017).

19. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M. & Eckner, R. A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. *Nature* **374**, 81–84 (1995).

20. Yang, X.-J., Ogryzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).

21. Eckner, R. *et al.* Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev.* **8**, 869–884 (1994).

22. Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO J* **28**, 948–958 (2009).

23. Bhaumik, P. *et al.* Structural insights into interactions of C/EBP transcriptional activators with the Taz2 domain of p300. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1914–1921 (2014).

24. Miller Jenkins, L. M. *et al.* Characterization of the p300 Taz2-p53 TAD2 complex and comparison with the p300 Taz2-p53 TAD1 complex. *Biochemistry* **54**, 2001–2010 (2015).

25. Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13260–13265 (2009).

26. Haberz, P., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding domain of CBP. *Protein Sci.* **25**, 2256–2267 (2016).

27. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).

28. Mori, S., Abeygunawardana, C., Johnson, M. O. & van Zijl, P. C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J Magn Reson B* **108**, 94–98 (1995).

29. Solyom, Z. *et al.* BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR* **55**, 311–321 (2013).

30. Bermel, W. *et al.* Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc.* **128**, 3918–3919 (2006).

31. Murrali, M. G., Piai, A., Bermel, W., Felli, I. C. & Pierattelli, R. Proline Fingerprint in Intrinsically Disordered Proteins. *Chembiochem* **19**, 1625–1629 (2018).

32. Bermel, W. *et al.* H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *J. Magn. Reson.* **198**, 275–281 (2009).

33. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).

34. Peng, J. W. & Wagner, G. [20] Investigation of protein motions via relaxation measurements. in *Methods in Enzymology* **239**, 563–596 (Academic Press, 1994).

35. Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).

36. Varadi, M., Guharoy, M., Zsolyomi, F. & Tompa, P. DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. *BMC Bioinformatics* **16**, 153 (2015).

37. Garrett, D. S., Seok, Y.-J., Peterkofsky, A., Clore, G. M. & Gronenborn, A. M. Identification by NMR of the Binding Surface for the Histidine-Containing Phosphocarrier Protein HPr on the N-Terminal Domain of Enzyme I of the Escherichia coli Phosphotransferase System. *Biochemistry* **36**, 4393–4398 (1997).

38. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1–16 (2013).

39. Felli, I. C. & Pierattelli, R. Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* **64**, 473–481 (2012).

40. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M. & Pierattelli, R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 25–45 (2006).

41. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*. (Springer International Publishing, 2015).

42. Pelka, P. *et al.* Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP. *Nucleic Acids Res.* **37**, 1095–1106 (2009).

43. Schöler, H. R., Ciesiolka, T. & Gruss, P. A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. *Cell* **66**, 291–304 (1991).

44. Agoff, S. N. & Wu, B. CBF mediates adenovirus Ela trans-activation by interaction at the C-terminal promoter targeting domain of conserved region 3. *Oncogene* **9**, 3707–3711 (1994).

45. Liu, F. & Green, M. R. Promoter targeting by adenovirus E1a through interaction with different cellular DNA-binding domains. *Nature* **368**, 520–525 (1994).

46. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **114**, 6561–6588 (2014).

47. Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956 (2007).
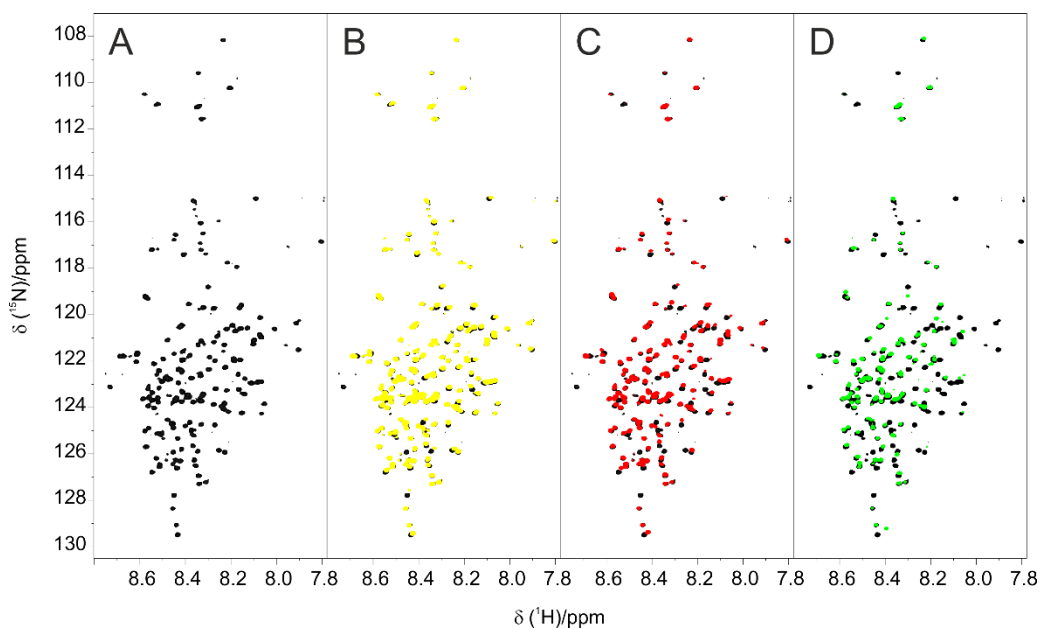
Figure 1. Titration of E1A12S with CBP-ID4 followed by ^{1}H-^{15}N BEST-TROSY experiments. A) E1A12S (black); B) E1A12S:CBP-ID4 at 1:0.5 molar ration (yellow); C) E1A12S:CBP-ID4 at 1:1 molar ratio (red); and D) E1A12S:CBP-ID4 at 1:2 molar ratio (green). All the experiments were acquired at 283 K, using a 22.3 T Bruker Avance III spectrometer equipped with a TCI CryoProbe^{TM}.
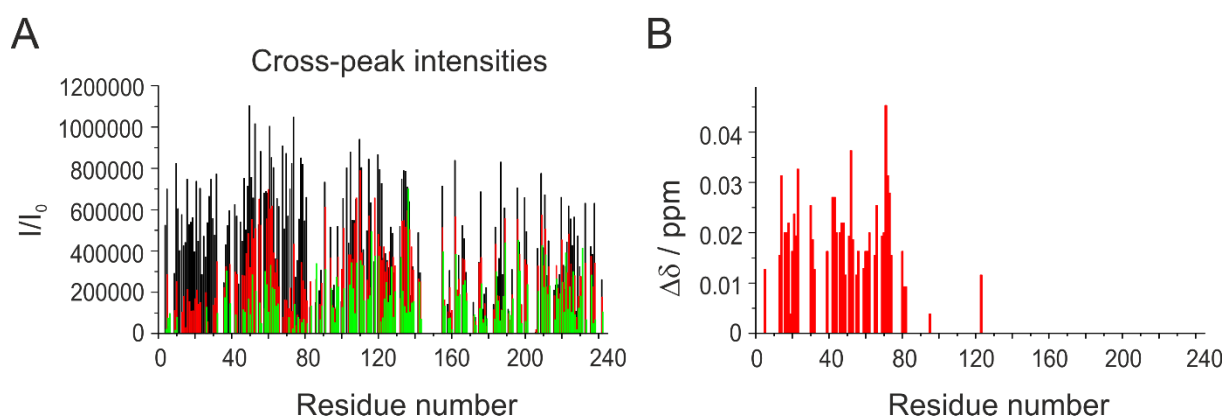

Figure 2. A) Comparison of cross-peak intensities in ^{1}H-^{15}N BEST-TROSY experiments for E1A12S in isolated form (black), E1A12S:ID4 1:1 complex (red) and E1A12S:ID4 1:2 complex (green). B) Mapping of the chemical shift perturbations of E1A12S signals in the E1A12S: ID4 1:1 complex using the Garrett equation.
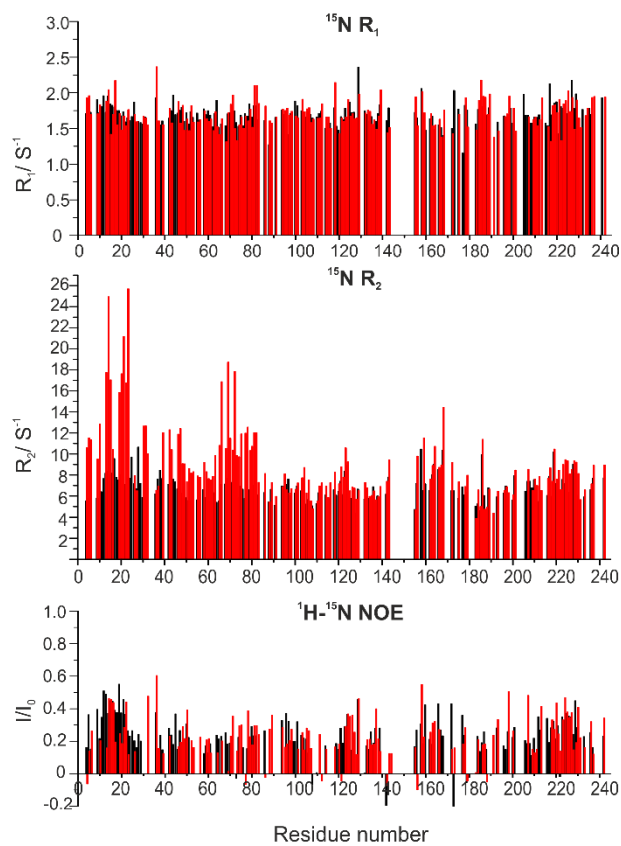
Figure 3. Nuclear relaxation data for backbone amide ^{15}N obtained for E1A12S (black) and for the 1:1 complex E1A12S:ID4 (red). From top to the bottom, ^{15}N longitudinal relaxation rates (R_1, Hz), ^{15}N transverse relaxation rates (R_2, Hz) and ^{1}H-^{15}N NOE.
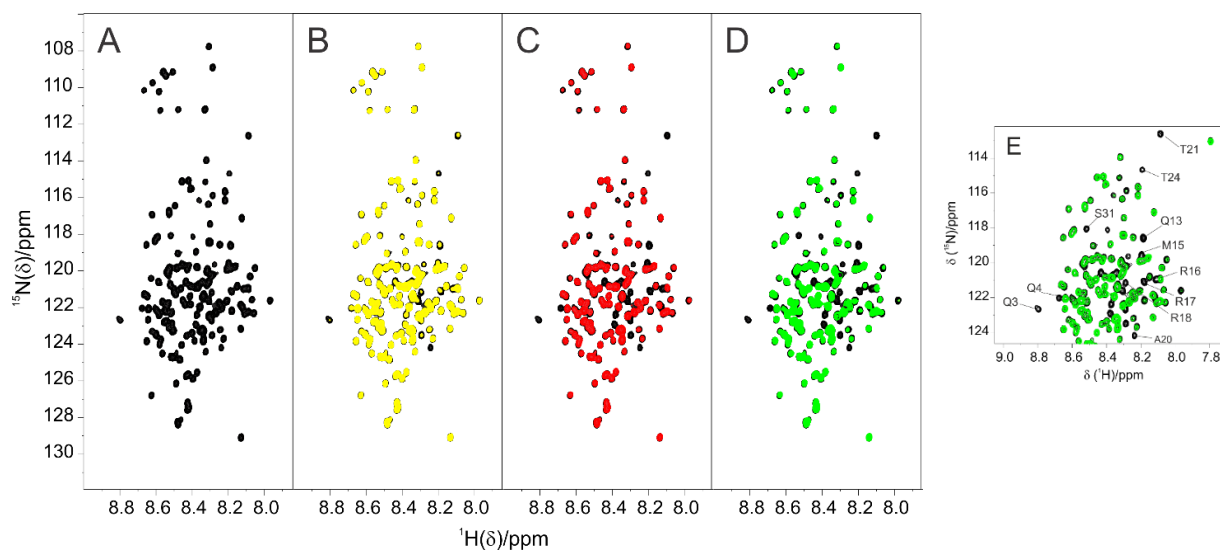
Figure 4. Titration of CBP-ID4 with E1A12S followed by ^{1}H-^{15}N FHSQC experiments A) CBP-ID4; B) CBP-ID4:E1A12S at 1:0.5 molar ratio (yellow); C) CBP-ID4:E1A12S at 1:1 molar ratio (red); D) CBP-ID4:E1A12S at 1:2 molar ratio (green) acquired at 283 K 22.3 T, using a Bruker Avance III spectrometer equipped with a TCI CryoProbe^{TM}. E) Comparison of a selected region of ^{1}H-^{15}N FHSQC spectrum of ID4 (black) and CBP-ID4:E1A12S 1:2 (green) showing the assignment of the most perturbed signals.
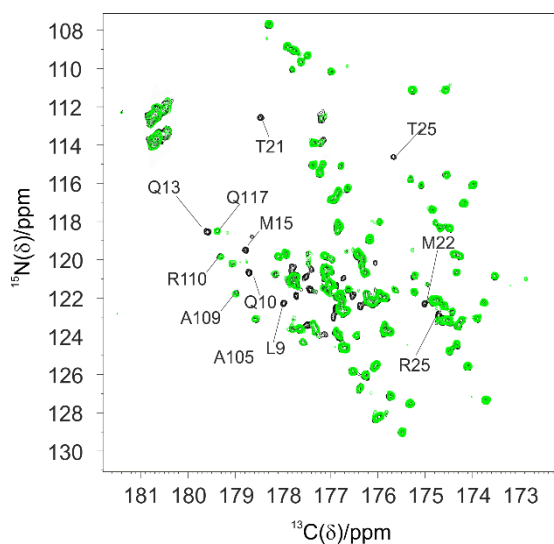

Figure 5. Comparison 2D (H^N)CON spectra of CBP-ID4 (black) and CBP-ID4:E1A12S 1:2 (green) acquired at 283 K with a 16.4 T Bruker Avance NEO spectrometer equipped with a ^{13}C TXO CryoProbe^{TM}. The assignment of selected peaks is also reported.
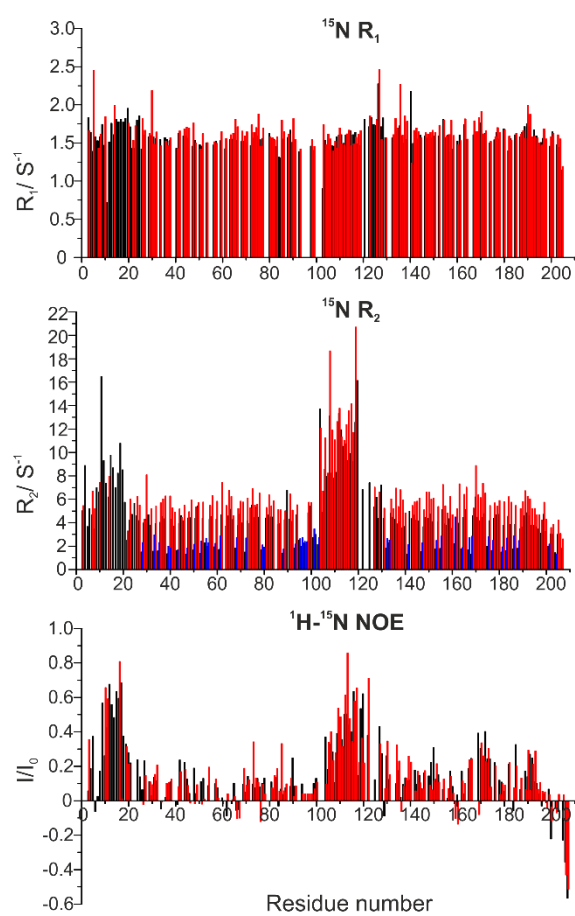
Figure 6. Nuclear relaxation data for backbone amide ^{15}N obtained for CBP-ID4 (black) and for the 1:1 complex CBP-ID4:E1A12S (red). From top to the bottom, ^{15}N longitudinal relaxation rates (R_1, Hz), ^{15}N transverse relaxation rates (R_2, Hz) including proline residues, and ^{1}H-^{15}N NOE. The relaxation rates R_2 of proline residues are represented in blue.
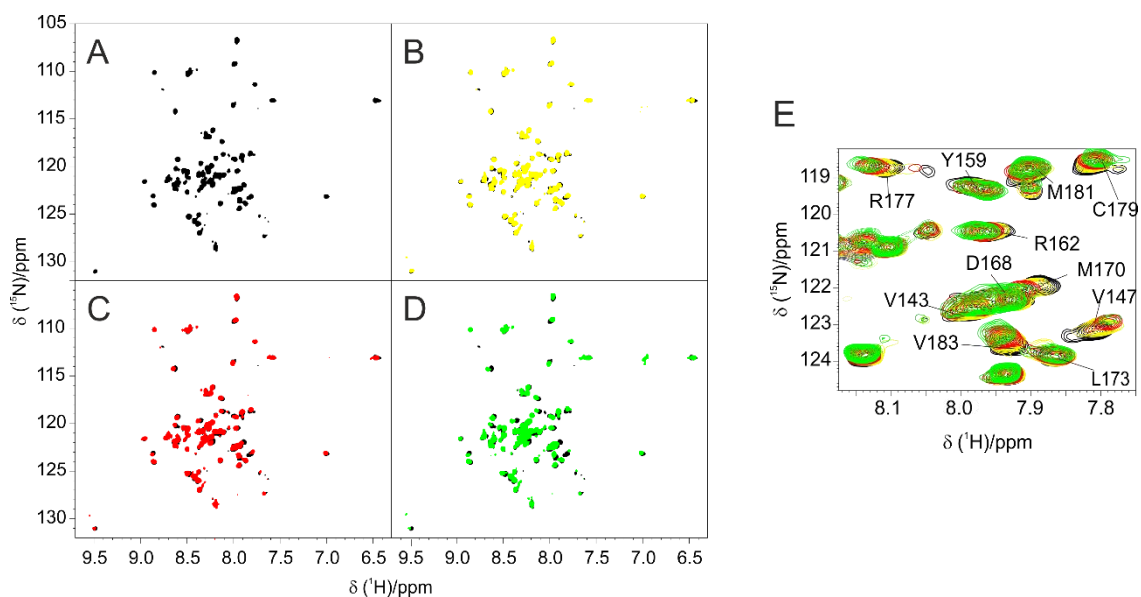
Figure 7. Titration of E1ACR3 with CBP-ID4 followed by ^{1}H-^{15}N HSQC experiments A) E1ACR3 (black); B) E1ACR3:
CBP-ID4 at 1:0.5 molar ratio (yellow); C) E1ACR3:ID4 at 1:1 molar ratio (red); D) E1ACR3:ID4 at 1:2 molar ratio
(green) acquired at 283 K , using a 22.3 T Bruker Avance III spectrometer equipped with a TCI CryoProbe^{TM}. E)
Selected region of the ^{1}H-^{15}N HSQC maps recorded at the various concentration (colour code as above).
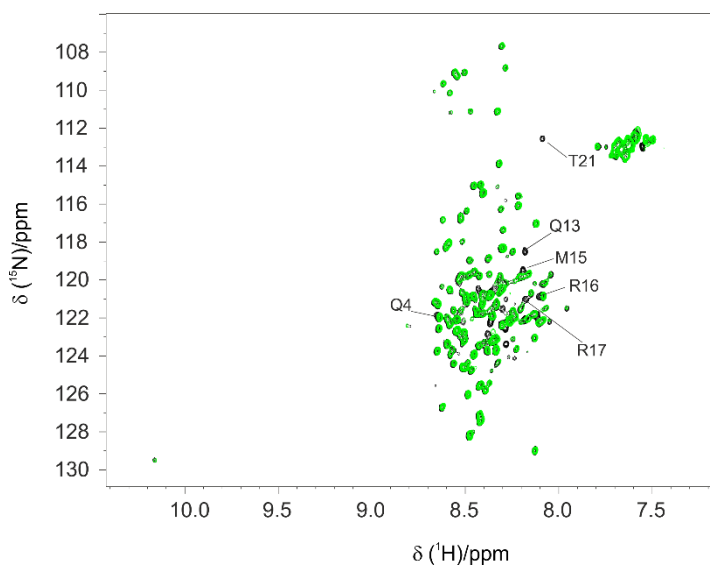

Figure 8. Comparison ^{1}H-^{15}N FHSQC spectra of CBP-ID4 (black) and CBP-ID4:E1ACR3 1:2 (green) acquired at 283
K with a 22.3 T Bruker Avance III spectrometer equipped with a TCI CryoProbe^{TM}. The assignment of selected
peaks is also reported.

Supplementary figures
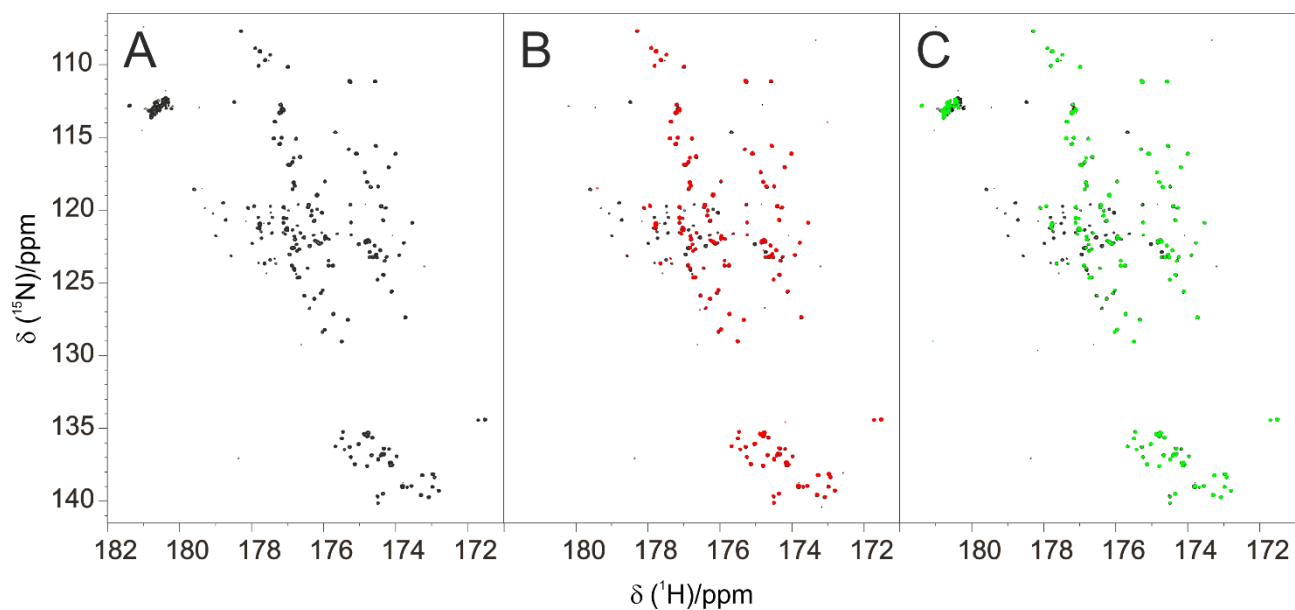
Figure S1. Titration of CBP-ID4 with E1A12S followed by 2D CON experiments. A) CBP-ID4 (black); B) CBP-ID4:E1A12S 1:1 (red); C) CBP-ID4:E1A12S 1:2 (green) acquired at 283 K, using a 16.4 T Bruker Avance NEO spectrometer equipped with a ^{13}C TXO CryoProbe^{TM}.
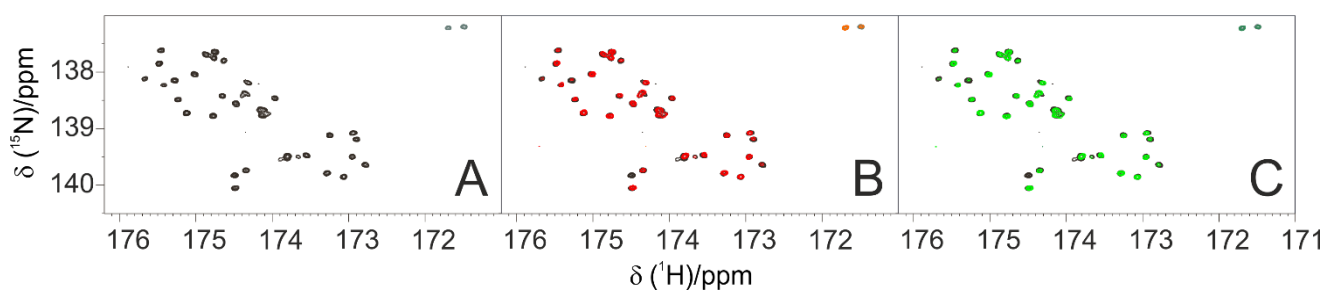

Figure S2. Expansion of the proline region in 2D CON spectra. A) CBP-ID4 (black); B) CBP-ID4:E1A12S 1:1 (red); C) CBP-ID4:E1A12S 1:2 (green) acquired at 283 K, using a 16.4 T Bruker Avance NEO spectrometer equipped with a ^{13}C TXO CryoProbe^{TM}.

Chapter 4. Conclusions

Intrinsically disordered proteins (IDPs) have nowadays an established role in protein science. Their abundance, their involvement in many key regulatory and signalling processes in cells, and their implication in several diseases enhance the interest in their investigation, stimulating novel discoveries about the roles of protein disorder in health and disease. Understanding at the molecular level all implications of protein intrinsic disorder, which has been overlooked until a few years ago, is going to have a tremendous impact in biomedicine and drug discovery.

NMR spectroscopy offers effective tools for the study of the structural and dynamical properties of disordered states of proteins at atomic resolution in systems as complex as whole cells. However, NMR experiments optimized for the study of folded proteins are not necessarily also optimal for IDPs. This has stimulated the development of novel approaches for their characterization.

The use of ^{13}C detected experiments in particular has proven to be an excellent strategy to investigate disordered proteins, providing high resolution information even at physiological condition. The first part of this doctoral thesis was then focused on the development of new methods that comprise different aspects of the structural characterization of IDPs. All these efforts to develop new NMR methods aimed to an improved knowledge and understanding of IDPs and their fundamental role in biological functions.

Thanks to the combination of NMR spectroscopy with other biophysical techniques it is possible to obtain information about the structure and the dynamics of disordered proteins, giving access to atomic resolution information for viral proteins responsible of very severe infections as in the case of Nipah virus. NMR spectroscopy can be exploit to unveil the role of disorder in promoting

phase transitions as liquid-liquid phase separation and how this phenomenon is involved in human pathologies. Using NMR it is possible to study the protein's interaction with small molecules, for example to test the activity of new drugs, unveiling the mechanism of action, an essential information to optimize the drug design. This is the case of nordihydroguaiaretic acid (NDGA), a phenolic dibenzenediol lignan molecular that prevents α-synuclein aggregation.

Considering the large abundance of IDPs, the occurrence of interactions between disordered proteins is significantly frequent but still poorly understood. Thanks to NMR spectroscopy it is possible to unveil the formation of flexible complexes between two IDPs, that may be involved in important biological function. This is the case of the viral replication of Adenovirus through the interaction of one of the proteins early expressed by the viral genome, namely the E1A protein, with the human CBP transcription coactivator.

The results obtained in this doctoral work provide novel examples to increase our knowledge on the biological role of intrinsic disorder in nature. NMR spectroscopy is recognized as a powerful technique for structural biology, particularly useful for the investigation of very dynamic systems. As new challenges are continuously appearing in Science, the development of new methods of investigation is essential, providing also new strategies to answer at least some of the endless biological questions still unanswered.

Bibliography

1. Creighton, T. E. *Proteins: structures and molecular properties*. (Macmillan, 1993).

2. Kulkarni, P. & Uversky, V. N. Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules* **9**, (2019).

3. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666 (1958).

4. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

5. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Bioinforma.* **41**, 415–427 (2000).

6. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).

7. Radivojac, P. *et al.* Intrinsic disorder and functional proteomics. *Biophys. J.* **92**, 1439–1456 (2007).

8. Van Der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).

9. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **114**, 6561–6588 (2014).

10. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).

11. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).

12. Tompa, P., Fersht, A. & Fersht, A. *Structure and Function of Intrinsically Disordered Proteins*. (Chapman and Hall/CRC, 2009).

13. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9 Suppl 2**, S1 (2008).

14. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta BBA-Proteins Proteomics* **1804**, 1231–1264 (2010).

15. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397–412 (1991).

16. Jensen, M. R., Zweckstetter, M., Huang, J. & Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).

17. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **579**, 3346–3354 (2005).

18. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets: the roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148 (2005).

19. Tompa, P. & Kovacs, D. Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol. Biochim. Biol. Cell.* **88**, 167–174 (2010).

20. Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).

21. Darling, A. L. & Uversky, V. N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front. Genet.* **9**, (2018).

22. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).

23. Uversky, V. N. Intrinsic disorder-based protein interactions and their modulators. *Curr. Pharm. Des.* **19**, 4191–4213 (2013).

24. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).

25. Contreras-Martos, S. *et al.* Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci. Rep.* **7**, 4676 (2017).

26. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).

27. Wu, S. *et al.* The Dynamic Multisite Interactions between Two Intrinsically Disordered Proteins. *Angew. Chem. Int. Ed Engl.* **56**, 7515–7519 (2017).

28. Van Roey, K. *et al.* Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **114**, 6733–6778 (2014).

29. Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956 (2007).

30. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).

31. Haynes, C. *et al.* Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLOS Comput. Biol.* **2**, e100 (2006).

32. Uversky, V. N. Intrinsic Disorder, Protein-Protein Interactions, and Disease. *Adv. Protein Chem. Struct. Biol.* **110**, 85–121 (2018).

33. Hashimoto, M., Rockenstein, E., Crews, L. & Masliah, E. Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *NeuroMolecular Med.* **4**, 21–35 (2003).

34. Masters, C. L. *et al.* Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. *EMBO J.* **4**, 2757–2763 (1985).

35. Uversky, V. N. Neuropathology, biochemistry, and biophysics of α-synuclein aggregation. *J. Neurochem.* **103**, 17–37 (2007).

36. Selkoe, D. J. Cell biology of protein misfolding: The examples of Alzheimer's and Parkinson's diseases. *Nat. Cell Biol.* **6**, 1054–1061 (2004).

37. Wells, M. *et al.* Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5762–5767 (2008).

38. Bellotti, V., Mangione, P. & Stoppini, M. Biological activity and pathological implications of misfolded proteins. *Cell. Mol. Life Sci. CMLS* **55**, 977–991 (1999).

39. Xue, B. *et al.* Structural disorder in viral proteins. *Chem. Rev.* **114**, 6880–6911 (2014).

40. Chinnadurai, G. Opposing oncogenic activities of small DNA tumor virus transforming proteins. *Trends Microbiol.* **19**, 174–183 (2011).

41. Xue, B. *et al.* Viral Disorder or Disordered Viruses: Do Viral Proteins Possess Unique Features? (2010).

42. Dyson, H. J. & Wright, P. E. How Do Intrinsically Disordered Viral Proteins Hijack the Cell? *Biochemistry* **57**, 4045–4046 (2018).

43. Butel, J. S. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. *Carcinogenesis* **21**, 405–426 (2000).

44. Boulanger, P. A. & Blair, G. E. Expression and interactions of human adenovirus oncoproteins. *Biochem. J.* **275**, 281–299 (1991).

45. Uversky, V. N., Roman, A., Oldfield, C. J. & Dunker, A. K. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J. Proteome Res.* **5**, 1829–1842 (2006).

46. DeCaprio, J. A. *et al.* SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. *Cell* **54**, 275–283 (1988).

47. Pelka, P., Ablack, J. N. G., Fonseca, G. J., Yousef, A. F. & Mymryk, J. S. Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes. *J. Virol.* **82**, 7252–7263 (2008).

48. Bernstein, F. C. *et al.* The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542 (1977).

49. Demers, J.-P., Chevelkov, V. & Lange, A. Progress in correlation spectroscopy at ultra-fast magic-angle spinning: basic building blocks and

complex experiments for the study of protein structure and dynamics. *Solid State Nucl. Magn. Reson.* **40**, 101–113 (2011).

50. Nannenga, B. L. & Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat. Methods* **16**, 369–379 (2019).

51. Kataoka, M. & Goto, Y. X-ray solution scattering studies of protein folding. *Fold. Des.* **1**, R107–R114 (1996).

52. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* **241**, 74–85 (2014).

53. Takeuchi, K., Baskaran, K. & Arthanari, H. Structure determination using solution NMR: Is it worth the effort? *J. Magn. Reson.* **306**, 195–201 (2019).

54. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).

55. Wishart, D. S. & Sykes, B. D. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363–392 (1994).

56. Meyer, B. & Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed.* **42**, 864–890 (2003).

57. Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **7**, 740–743 (2000).

58. Felli, I. C. & Pierattelli, R. Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* **64**, 473–481 (2012).

59. Brutscher, B. *et al.* NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv. Exp. Med. Biol.* **870**, 49–122 (2015).

60. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17**, 3–14 (2007).

61. Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **19**, 23–30 (2009).

62. Kovacs, H., Moskau, D. & Spraul, M. Cryogenically cooled probes—a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**, 131–155 (2005).

63. Serber, Z. *et al.* New Carbon-Detected Protein NMR Experiments Using CryoProbes. *J. Am. Chem. Soc.* **122**, 3554–3555 (2000).

64. Wirmer, J., Berk, H., Ugolini, R., Redfield, C. & Schwalbe, H. Characterization of the unfolded state of bovine alpha-lactalbumin and comparison with unfolded states of homologous proteins. *Protein Sci. Publ. Protein Soc.* **15**, 1397–1407 (2006).

65. Bermel, W. *et al.* High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J. Biomol. NMR* **57**, 353–361 (2013).

66. Bermel, W. *et al.* Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). *Chembiochem Eur. J. Chem. Biol.* **13**, 2425–2432 (2012).

67. Bertini, I., Felli, I. C., Gonnelli, L., Kumar M V, V. & Pierattelli, R. [13]C direct-detection biomolecular NMR spectroscopy in living cells. *Angew. Chem. Int. Ed Engl.* **50**, 2339–2341 (2011).

68. Bermel, W., Bertini, I., Felli, I. C., Peruzzini, R. & Pierattelli, R. Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins. *ChemPhysChem Eur. J. Chem. Phys. Phys. Chem.* **11**, 689–695 (2010).

69. Felli, I. C. & Pierattelli, R. Novel methods based on ^{13}C detection to study intrinsically disordered proteins. *J. Magn. Reson.* **241**, 115–125 (2014).

70. Cook, E. C., Usher, G. A. & Showalter, S. A. The Use of ^{13}C Direct-Detect NMR to Characterize Flexible and Disordered Proteins. *Methods Enzymol* **611**, 81–100 (2018).

71. Shimba, N. *et al.* Optimization of ^{13}C direct detection NMR methods. *J. Biomol. NMR* **30**, 175–179 (2004).

72. Wong, L. E., Maier, J., Wienands, J., Becker, S. & Griesinger, C. Sensitivity-Enhanced Four-Dimensional Amide-Amide Correlation NMR Experiments for Sequential Assignment of Proline-Rich Disordered Proteins. *J. Am. Chem. Soc.* **140**, 3518–3522 (2018).

73. Motáčková, V. *et al.* Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J. Biomol. NMR* **48**, 169–177 (2010).

74. Kazimierczuk, K., Stanek, J., Zawadzka-Kazimierczuk, A. & Koźmiński, W. High-Dimensional NMR Spectra for Structural Studies of Biomolecules. *ChemPhysChem* **14**, 3015–3025 (2013).

75. Kupče, Ē. & Freeman, R. Projection−Reconstruction Technique for Speeding up Multidimensional NMR Spectroscopy. *J. Am. Chem. Soc.* **126**, 6429–6440 (2004).

76. Kazimierczuk, K., Zawadzka, A., Koźmiński, W. & Zhukov, I. Random sampling of evolution time space and Fourier transform processing. *J. Biomol. NMR* **36**, 157–168 (2006).

77. Kazimierczuk, K., Zawadzka, A., Koźmiński, W. & Zhukov, I. Lineshapes and artifacts in Multidimensional Fourier Transform of arbitrary sampled NMR data sets. *J. Magn. Reson.* **188**, 344–356 (2007).

78. Kazimierczuk, K. & Orekhov, V. Non-uniform sampling: post-Fourier era of NMR data collection and processing. *Magn. Reson. Chem.* **53**, 921–926 (2015).

79. Hiller, S., Fiorito, F., Wüthrich, K. & Wider, G. Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci.* **102**, 10876–10881 (2005).

80. Fiorito, F., Hiller, S., Wider, G. & Wüthrich, K. Automated resonance assignment of proteins: 6D APSY-NMR. *J. Biomol. NMR* **35**, 27–37 (2006).

81. Gossert, A. D., Hiller, S. & Fernández, C. Automated NMR resonance assignment of large proteins for protein-ligand interaction studies. *J. Am. Chem. Soc.* **133**, 210–213 (2011).

82. Felli, I. C. & Pierattelli, R. Spin-state-selective methods in solution- and solid-state biomolecular 13C NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **84–85**, 1–13 (2015).

83. Nováček, J. *et al.* 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J. Biomol. NMR* **50**, 1–11 (2011).

84. Nováček, J., Haba, N. Y., Chill, J. H., Zídek, L. & Sklenář, V. 4D non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J. Biomol. NMR* **53**, 139–148 (2012).

85. Hošek, T., Gil-Caballero, S., Pierattelli, R., Brutscher, B. & Felli, I. C. Longitudinal relaxation properties of (1)H(N) and (1)H(α) determined by direct-detected (13)C NMR experiments to study intrinsically disordered proteins (IDPs). *J. Magn. Reson.* **254**, 19–26 (2015).

86. Kupče, E. NMR with multiple receivers. *Top. Curr. Chem.* **335**, 71–96 (2013).

87. Kupče, Ē., Freeman, R. & John, B. K. Parallel Acquisition of Two-Dimensional NMR Spectra of Several Nuclear Species. *J. Am. Chem. Soc.* **128**, 9606–9607 (2006).

88. Viegas, A. *et al.* UTOPIA NMR: activating unexploited magnetization using interleaved low-gamma detection. *J. Biomol. NMR* **64**, 9–15 (2016).

89. Gallo, A., Franks, W. T. & Lewandowski, J. R. A suite of solid-state NMR experiments to utilize orphaned magnetization for assignment of proteins using parallel high and low gamma detection. *J. Magn. Reson.* **305**, 219–231 (2019).

90. Kupče, E. & Kay, L. E. Parallel acquisition of multi-dimensional spectra in protein NMR. *J. Biomol. NMR* **54**, 1–7 (2012).

91. Chakraborty, S., Paul, S. & Hosur, R. V. Simultaneous acquisition of ^{13}Cα-^{15}N and ^{1}H-^{15}N-^{15}N sequential correlations in proteins: application of dual receivers in 3D HNN. *J. Biomol. NMR* **52**, 5–10 (2012).

92. Romero, P. *et al.* Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* 437–448 (1998).

93. Ottolini, D., Calí, T., Szabò, I. & Brini, M. Alpha-synuclein at the intracellular and the extracellular side: functional and dysfunctional implications. *Biol. Chem.* **398**, 77–100 (2016).

94. Xu, J. *et al.* Dopamine-dependent neurotoxicity of α-synuclein: A mechanism for selective neurodegeneration in Parkinson disease. *Nat. Med.* **8**, 600–606 (2002).

95. Masliah, E. *et al.* Dopaminergic Loss and Inclusion Body Formation in α-Synuclein Mice: Implications for Neurodegenerative Disorders. *Science* **287**, 1265–1269 (2000).

96. Olanow, C. W., Obeso, J. A. & Stocchi, F. Drug Insight: continuous dopaminergic stimulation in the treatment of Parkinson's disease. *Nat. Clin. Pract. Neurol.* **2**, 382–392 (2006).

97. Pellecchia, M., Sem, D. S. & Wüthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.* **1**, 211–219 (2002).

98. Ambadipudi, S. & Zweckstetter, M. Targeting intrinsically disordered proteins in rational drug discovery. *Expert Opin. Drug Discov.* **11**, 65–77 (2016).

99. Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* **14**, 1051–1057 (2009).

100. Tamiola, K., Acar, B. & Mulder, F. A. A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **132**, 18000–18003 (2010).

101. Tamiola, K. & Mulder, F. A. A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40**, 1014–1020 (2012).

102. Uversky, V. N. Protein intrinsic disorder-based liquid–liquid phase transitions in biological systems: Complex coacervates and membrane-less organelles. *Adv. Colloid Interface Sci.* **239**, 97–114 (2017).

103. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).

104. Posey, A. E., Holehouse, A. S. & Pappu, R. V. Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol* **611**, 1–30 (2018).

105. Wegmann, S. *et al.* Tau protein liquid–liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, e98049 (2018).

106. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).

107. Peskett, T. R. *et al.* A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation. *Mol. Cell* **70**, 588-601.e6 (2018).

108. Wang, W. & Wang, D. Extreme Fuzziness: Direct Interactions between Two IDPs. *Biomolecules* **9**, 81 (2019).

109. Hošek, T. *et al.* Structural and Dynamic Characterization of the Molecular Hub Early Region 1A (E1A) from Human Adenovirus. *Chemistry* **22**, 13010–13013 (2016).

110. De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. *J. Mol. Biol.* **303**, 243–253 (2000).

111. De Guzman, R. N., Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* **44**, 490–497 (2005).

112. Brüschweiler, S., Konrat, R. & Tollinger, M. Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem. Biol.* **8**, 1600–1610 (2013).

113. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521–534 (2004).

114. Kjaergaard, M., Andersen, L., Nielsen, L. D. & Teilum, K. A folded excited state of ligand-free nuclear coactivator binding domain (NCBD) underlies plasticity in ligand recognition. *Biochemistry* **52**, 1686–1693 (2013).

115. Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13260–13265 (2009).

116. Haberz, P., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding domain of CBP. *Protein Sci. Publ. Protein Soc.* **25**, 2256–2267 (2016).

117. Piai, A. *et al.* Just a Flexible Linker? The Structural and Dynamic Properties of CBP-ID4 Revealed by NMR Spectroscopy. *Biophys. J.* **110**, 372–381 (2016).

118. Olsen, J. G., Teilum, K. & Kragelund, B. B. Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **74**, 3175–3183 (2017).

119. Bonvin, A. M., Boelens, R. & Kaptein, R. NMR analysis of protein interactions. *Curr. Opin. Chem. Biol.* **9**, 501–508 (2005).

120. Ortega-Roldan, J. L., Blackledge, M. & Jensen, M. R. Characterizing Protein-Protein Interactions Using Solution NMR Spectroscopy. in *Protein Complex Assembly: Methods and Protocols* (ed. Marsh, J. A.) 73–85 (Springer New York, 2018).